DANIELA LUCAS DA SILVA

UMA PROPOSTA METODOLÓGICA

PARA CONSTRUÇÃO DE ONTOLOGIAS:

UMA PERSPECTIVA INTERDISCIPLINAR

ENTRE AS CIÊNCIAS DA INFORMAÇÃO E DA COMPUTAÇÃO

Belo Horizonte

Escola de Ciência da Informação da UFMG

2008

DANIELA LUCAS DA SILVA

UMA PROPOSTA METODOLÓGICA

PARA CONSTRUÇÃO DE ONTOLOGIAS:

UMA PERSPECTIVA INTERDISCIPLINAR

ENTRE AS CIÊNCIAS DA INFORMAÇÃO E DA COMPUTAÇÃO

Dissertação apresentada ao programa de pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Área de Concentração: Organização e Uso da Informação.

Orientador: Prof. Dr. Renato Rocha Souza
Co-Orientador: Prof.Dr. Maurício Barcellos Almeida

Belo Horizonte

Escola de Ciência da Informação da UFMG

2008

DEDICATÓRIA

Primeiramente a Deus pelas sábias orientações nos caminhos árduos da vida, os quais, com paciência, são convertidos em paz.

À minha mãe, Fátima, pelo amor, incentivo e dedicação inigualáveis.

Ao meu pai, Ricardo, pelo amor, olhar e cuidado.

Aos meus irmãos, Marcelo e Leandro, pela força e incentivos constantes.

Ao Alan, pela paciência, apoio e carinho ao final dessa caminhada.

Aos amigos, pela compreensão na ausência e pela torcida.

AGRADECIMENTOS

Ao Prof. Renato, pela amizade e pela valiosa orientação e contribuição intelectual neste percurso acadêmico.

Ao Prof. Maurício, pelo apoio constante nas orientações sobre ontologias, sem as quais esse trabalho não teria sido concretizado.

Aos Profs. Maurício Barcellos Almeida, Lídia Alvarenga, Maria Aparecida Moura e Manoel Palhares, pelas preciosas observações no processo de qualificação.

A todos os professores da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, que souberam demonstrar a importância das pesquisas em Ciência da Informação para a sociedade.

A todos os colegas e funcionários do curso de mestrado e doutorado em Ciência da Informação da UFMG.

O sentimento cria a idéia.
A idéia gera o desejo.
O desejo acalentado forma a palavra.
A palavra orienta a ação.
A ação detona resultados.
Os resultados nos traçam o caminho nas áreas infinitas do tempo.
Cada criatura permanece na estrada que construiu para si mesma.
A escolha é sempre nossa.

EMMANUEL

RESUMO

A pesquisa apresenta princípios metodológicos para desenvolvimento de ontologias formais, construídos através da análise da literatura sobre metodologias para construção de ontologias e de normas internacionais para construção de *software* e construção de vocabulários controlados. Através de pesquisa teórica e empírica, foi possível construir um arcabouço metodológico interdisciplinar para o processo de construção de ontologias de domínio em especial e de outros tipos de ontologias, integrando princípios teóricos e metodológicos da Ciência da Informação, da Ciência da Computação, além de contribuições de diversas metodologias e métodos conhecidos para construção de ontologias.

Palavras-chave: representação do conhecimento, ontologias, construção de ontologias.

ABSTRACT

The research shows methodological principles for the development of formal ontologies, supported by the analysis of a literature about methodologies for building ontologies and the international standards for software engineering and controlled vocabularies architecture. Through theoretical and empiric researches, it was possible to form an interdisciplinary methodological framework for the process of building domain ontologies in particular and other types of ontology, integrating theoretical and methodological principles from Information Science, from Computer Science, besides other contributions of several methodologies and used methods for building ontologies.

Key-words: knowledge representation, ontologies, ontologies building

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1- Introdução

A organização da informação tornou-se um processo fundamental na medida em que vêm crescendo exponencialmente o volume de informações disponível, resultando muitas vezes na desorganização de acervos informacionais e conseqüentemente na dificuldade de se encontrar o que se procura num determinado sistema de recuperação de informação.

Com o advento das novas tecnologias da informação, houve um aumento considerável na produção e intercâmbio de registros em diversas áreas do conhecimento. Diante desse fato, a preocupação em organizar, armazenar e recuperar informações torna-se um grande desafio por parte de diversas classes profissionais, tais como bibliotecários, administradores, engenheiros, médicos, dentre outras. Tal desafio visa a atender às necessidades informacionais dessas comunidades de usuários, que exigem informações precisas na condução de suas atividades.

Nesse sentido, pesquisas têm sido desenvolvidas progressivamente visando ao desenvolvimento de mecanismos de indexação, organização e recuperação de informações, com o objetivo único de melhorar a eficácia dos sistemas de recuperação de informação. Podemos citar, dentre outras, algumas pesquisas nessa perspectiva voltadas à exploração semântica da informação, tais como: a) a Web Semântica que pretende criar metodologias, tecnologias e padrões de metadados para aumentar o escopo das atividades desempenhadas automaticamente (BERNERS-LEE, HENDLER e LASSILA, 2001); b) a utilização semântica embutida nos próprios documentos com o uso de estruturas da linguagem natural como os sintagmas nominais e verbais (SOUZA, 2005); c) instrumentos de representação de relacionamentos semânticos e conceituais como as ontologias (GRUBER, 1993) e os tesauros (DAHLBERG, 1978), objetivando evitar problemas relacionados à ambigüidade inerente às palavras da linguagem natural.

No campo da Ciência da Informação (CI), uma área que vem chamando a atenção para pesquisas, no sentido de composição de teorias e métodos visando à concepção de instrumentos voltados a organização e recuperação de informações, é a Inteligência Artificial (IA). Para Saracevic (1996), a IA divide-se em IA fraca e IA forte. A primeira vislumbra as tecnologias envolvidas na construção de sistemas de informações. Já a segunda preocupa-se com aspectos filosóficos do conhecimento representado na mente ou no computador. Para a Ciência da Informação, tanto a IA fraca quanto a forte tem interesse direto. Segundo Saracevic (1996, p.52),

> A IA fraca é fonte de muitas das inovações nos sistemas de informação, tais como sistemas inteligentes, hipertextos, bases de conhecimento, interfaces inteligentes e as questões sobre a interação homem-computador - todas elas de interesse para a CI e para as quais ela pode contribuir diretamente. A IA forte é a fonte do modelo teórico da cognição, no qual a informação, enquanto fenômeno, desempenha o mais importante papel. Portanto, esse modelo pode também contribuir para a pesquisa básica em CI. (SARACEVIC, 1996, p.52)

Esse fato contribui para a atenção dada às ontologias, cuja origem se dá no campo teórico da Filosofia (CORAZZON, 2008), sendo ainda pesquisadas e desenvolvidas como instrumento de representação de conhecimento nos campos das Ciências da Computação e da Informação. Para a CI as ontologias são de interesse pela potencialidade que elas têm em organizar e representar informação. Segundo Almeida e Bax (2003), as ontologias podem melhorar os processos de recuperação de informação ao organizar o conteúdo de fontes de dados num determinado domínio.

O termo ontologia possui muitas definições que serão exploradas neste trabalho, mas por ora pode-se afirmar que são estruturas que congregam conjuntos de conceitos padronizados, termos e definições aceitas por uma comunidade particular, e a forma com que estes conceitos estão relacionados entre si. Falbo (1998) afirma que o emprego de ontologias pode facilitar a comunicação entre diversos usuários com visões diferentes do processo através de um vocabulário compartilhado do domínio de interesse.

Entretanto, um problema, do ponto de vista metodológico, é que não há um padrão para construção de ontologias (JONES, BENCH-CAPON e VISSER, 1998; FERNÁNDEZ et al., 1999; USCHOLD e GRUNINGER, 1996; FALBO, 1998). Apesar de uma grande quantidade de ontologias já ter sido desenvolvida por diferentes comunidades - na química (GÓMEZ-PEREZ, FERNANDEZ e VICENTE, 1996) e na modelagem de negócio (GRUNINGER e FOX, 1995), só para citar alguns exemplos - sob diferentes abordagens e usando diferentes métodos e técnicas, não há consenso sobre uma metodologia para o processo de construção (FERNANDEZ, GÓMEZ-PEREZ e JURISTO, 1997). A conseqüência, segundo Fernandez, Gomez-Perez e Juristo (1997), é a ausência de atividades padronizadas, cujas realizações são conduzidas de forma artesanal e não como uma atividade científica. Além disto, verifica-se a falta de explicação sistemática de como e onde serão usadas as abordagens teóricas dentro de seu processo de elaboração.

Tal questão nos leva a investigar princípios teóricos e metodológicos que dêem sustentação científica no processo de construção de ontologias. Alguns pesquisadores da Ciência da Informação (SOERGEL, 1997 e 1999; VICKERY, 1997) que investigam aplicabilidades e contribuições das ontologias na área, apontam esse instrumento como um

tipo de classificação por lidar com categorização de conceitos, principio básico da teoria da classificação. Tal princípio também é empregado na elaboração de instrumentos tradicionais utilizados na biblioteconomia para fins de indexação e recuperação de documentos como, por exemplo, os tesauros. Segundo Campos (2001), técnicas de organização e uso da informação, advindas das teorias da Ciência da Informação, podem ser empregadas na construção de instrumentos terminológicos como os tesauros, cuja contribuição seria: a) uma sistematização de conceitos para estruturação de um domínio de conhecimento, objetivando organizar efetivamente um conhecimento em uma área específica; b) representar as relações entre conceitos; c) revelar conceitos ainda inexistentes ou redundantes, possibilitando uma normalização da terminologia.

O presente trabalho se insere nesse contexto, na medida em que busca uma abordagem teórica e metodológica para construção de ontologias.

É válido ressaltar neste ponto que os termos *informação*, dado, documento, conteúdo e *conhecimento* são utilizados na Ciência da Informação muitas das vezes sem distinção. Isso pode ser conferido em uma passagem em Freire (2006, p.10):

> Mas, o que torna paradigmático o trabalho de Otlet e la Fontaine não é somente sua percepção da nova relevância do antigo fenômeno da informação, como resumiram Wersig e Neveling (1975). O principal é que, no processo de criação dos instrumentos necessários para sua organização, armazenagem e recuperação, Otlet e la Fontaine centraram seus esforços no conteúdo dos documentos, ou seja, na informação em si, e isto foi realmente inovador pois até então nunca havia sido feito.
>
> Assim, esse novo paradigma nasce da atenção dada ao conteúdo (informação, conhecimento registrado) dos documentos, mais do que aos próprios documentos, ao tempo que estabelece uma certa fronteira entre as bibliotecas anteriores e os centros de documentação, onde a resposta à procura de informações e documentos, por temas e outros novos critérios, torna-se possível e amplia-se de forma espetacular.

O presente trabalho não se ocupa em fazer distinções teóricas acerca dos termos, apesar de tal investigação ser objeto de estudo da Ciência da Informação (BUCKLAND, 1991; CORNELIUS, 2002; CAPURRO, 2003).

No campo da Ciência da Computação, o termo *conhecimento* é utilizado na área de Inteligência Artificial no desenvolvimento de sistemas baseados em conhecimento e sistemas especialistas. Para tal, utilizam-se formalismos para a representação das bases de conhecimentos (REZENDE, 2003; RUSSELL e NORVIG, 2004). Já os termos dado, informação e conteúdo são usados em áreas da computação que se ocupam do desenvolvimento e uso de tecnologias destinadas aos sistemas de informações como sistemas

de bancos de dados, ferramentas de modelagem, linguagens de marcação, dentre outras (PRESSMAN, 2002; BOOCH, JACOBSON e RUMBAUGH, 2006; BAX, 2001).

Nessa pesquisa, os termos *informação* e *conhecimento* são usados com maior freqüência. Portanto, compromete-se, aqui, em direcionar o entendimento acerca de *conhecimento* e *informação* em um nível que permita a ambos registrar e comunicar por meio de uma linguagem, seja ela verbal ou notacional.

1.1 Pressupostos e questão de pesquisa

Acredita-se que o aporte teórico e metodológico existente no âmbito da Ciência da Informação possa contribuir na proposição de metodologias para o processo de elaboração de ontologias. Além disso, acredita-se que técnicas advindas da Ciência da Computação, como as formas de representação baseadas em lógica (que possibilitam o uso de mecanismos de inferência) e a marcação semântica de dados (que permitem, através dos metadados, atribuir significados mais precisos à informação), possam ser uma evolução em relação às técnicas tradicionais de representação e recuperação de informação. Afora a Ciência da Computação e a Ciência da Informação, há possibilidade de obter contribuições advindas da filosofia, da terminologia, dentre outras áreas.

Pressupõe-se uma correlação entre organização e uso da informação e técnicas computacionais. Acredita-se que pesquisas em organização e uso da informação possam contribuir sobremaneira para o desenvolvimento de instrumentos tecnológicos de qualidade, a partir de uma perspectiva da organização da informação centrada no usuário. Defende-se nesta pesquisa que a fertilização cruzada de tecnologias e metodologias das áreas da Ciência da Informação e Ciência da Computação pode conduzir a melhores resultados. Essa percepção interdisciplinar motiva a investigação de um instrumento computacional, as ontologias, capaz de representar e organizar estruturas conceituais e relacionais de um domínio de conhecimento.

Diante ao exposto, a **questão de pesquisa** encontra-se centrada em confirmar a possibilidade de proposição de princípios metodológicos para construção de ontologias fundamentados nas Ciências da Informação e da Computação, dentre outros campos do conhecimento. Nesse sentido, a questão é formulada como se segue: *Será possível confirmar a possibilidade de proposição de princípios metodológicos para construção de ontologias, à luz das contribuições pertinentes das Ciências da Informação e da Computação, dentre outros campos do conhecimento?*

1.2 Objetivos

Como **objetivo geral**, essa pesquisa almeja *propor princípios metodológicos para a construção de ontologias de domínio para representação do conhecimento em sistemas de informações.*

A consecução deste objetivo se dará através da análise da literatura sobre metodologias para construção de ontologias e de normas internacionais para construção de *software* e construção de vocabulários controlados. Serão ainda analisados os aportes das Ciências da Informação, da Computação e campos correlatos, para a correta sustentação teórica deste objetivo.

Como **objetivos específicos**, pretende-se:

- Identificar metodologias relevantes existentes para construção de ontologias;
- Identificar metodologias de construção de vocabulários controlados;
- Analisar comparativamente os objetos empíricos escolhidos, de forma a subsidiar a proposta metodológica.

Esta dissertação está estruturada da seguinte forma:

Nesta **introdução** são apresentados a motivação do estudo, os pressupostos, a questão de pesquisa e os objetivos a serem alcançados.

No capítulo de **fundamentação teórica** são apresentadas as contribuições teóricas de campos-chave que tratam da temática Representação do Conhecimento, destacando a Ciência da Informação e a Ciência da Computação. Tais contribuições são necessárias ao correto entendimento da proposta de pesquisa.

No capítulo sobre a **metodologia de pesquisa** são apresentados os passos que foram adotados para o desenvolvimento da pesquisa de forma a atingir os objetivos do trabalho.

O capítulo de **análise dos objetos empíricos** abrange a análise e a apresentação das metodologias e dos métodos para construção de ontologias, bem como da metodologia e da norma de construção de vocabulários controlados, com vistas à consecução do objetivo geral.

No capítulo de **apresentação da análise comparativa** são analisados comparativamente os objetos empíricos, de modo a fazer um escopo comparativo, o qual servirá de subsídio para a proposta final deste trabalho.

Tal proposta, por conseguinte, é apresentada no capítulo **proposta metodológica para construção de ontologias**, em que são apresentadas as fases constituintes da proposta metodológica, em conjunto com os seus princípios teóricos e metodológicos. E, por fim, o capítulo de **conclusões e perspectivas futuras** é dedicado a reflexões qualitativas acerca dos resultados alcançados, englobando as contribuições para a Ciência da Informação, bem como a outros campos do conhecimento dedicados à Representação da Informação e do Conhecimento. Neste último capítulo também serão discutidos caminhos para futuros trabalhos que poderão ser desenvolvidos em decorrência do que aqui se apresenta.

2- Fundamentação Teórica

A representação de conhecimento tem sido tratada como objeto de estudos em diferentes campos do conhecimento. Esta pesquisa está concentrada em alguns desses campos, dentre os quais podemos destacar a Filosofia, a Terminologia, a Ciência da Informação e a Ciência da Computação.

A *Filosofia* é um campo preocupado em estudar as leis de raciocínio. A *lógica* se ocupa dessas leis e tem sido definida como a ciência das leis do pensamento (COPI, 1981). O estudo do raciocínio é um ponto em destaque nesta pesquisa, pois é a partir dele que se representam declarações sobre todos os tipos de coisas no mundo e sobre as relações entre elas. Nesse sentido, é possível deduzir ou inferir conhecimento a partir dessas declarações. Os *silogismos* são uma contribuição importante do filósofo Aristóteles às áreas de Inteligência Artificial e a Engenharia ontológica no sentido de ser um padrão formal para representar regras de inferências. Aristóteles também contribui ao determinar quais distinções possíveis poderiam ser direcionadas às coisas do mundo. Dessa forma, estabeleceu, através da *abstração*, categorias gerais (CORAZZON, 2008a) na ação de classificar as coisas, buscando uma organização do conhecimento a respeito do mundo. Tais contribuições serão apresentadas na seção 2.1.1.

A utilização de instrumentos de controle terminológico torna-se fundamental para os sistemas de informação automatizados, tendo em vista aprimorar o tratamento das informações e os processos de recuperação da informação e, consequentemente, melhorar a interação do usuário em suas estratégias de busca. A *Terminologia* concentra seus esforços na sistematização de conceitos para a elaboração de definições consistentes sobre eles e será apresentada na seção 2.1.2.

Na *Ciência da Informação*, a preocupação concentra-se na organização de conteúdos de documentos e na recuperação de informação através da elaboração de linguagens documentárias. No processo de elaboração de linguagens documentárias, a representação da informação nos sistemas de recuperação de informação tem fundamental importância, pois define a terminologia que será direcionada à base de dados de documentos, auxiliando o usuário na elaboração de estratégias de busca. O consenso em relação à terminologia a ser adotada pelo sistema é essencial para a comunicação entre usuário e sistema. Para tal, buscam-se os aportes teóricos da categorização e classificação de conceitos

(RANGANATHAN 1967; DAHLBERG, 1978) para representar o conhecimento do assunto. As contribuições desse campo serão apresentadas na seção 2.2.

No âmbito da *Ciência da Computação*, os estudos sobre concepção de modelos destinados à construção de sistemas têm sido cada vez mais intensificados na busca de uma representação mais próxima da realidade. Nesse sentido, destaca-se o modelo de entidade e relacionamento (CHEN, 1976) e o modelo orientado a objetos (BOOCH, JACOBSON e RUMBAUGH, 2006), além dos fundamentos da lógica aplicados na representação de bases de conhecimento (RUSSELL e NORVIG, 2004). Atualmente, uma área que vem se destacando na busca por uma organização e representação formal de conceitos e relações através de um compromisso ontológico é a Engenharia ontológica (GRUBER, 1993a; GUARINO, 1998), que será destaque nesta pesquisa. As contribuições da Ciência da Computação serão apresentadas na seção 2.3.

2.1 Considerações iniciais: Representação do Conhecimento

O conceito de representação pode ser definido como uma reprodução ou cópia de algum objeto ou fato. Quando falamos de representação de informação e de conhecimento, estamos tratando de um processo complexo, uma vez que estamos tentando reproduzir em algum suporte (papel, tela, dispositivo eletrônico), através de símbolos ou linguagens, conceitos e significados construídos ou percebidos pela mente humana. Para tal, deve haver alguma aproximação entre o substituto e seu referente planejado no mundo. Esta correspondência é a **semântica** da representação. Campos (2004, p.24) assinala que a representação perfeita é, em geral, impossível, tanto na prática quanto em princípio. A única representação completamente precisa de um objeto é o objeto em si. Neste sentido, a representação é, fundamentalmente, um substituto para aquilo que representa.

Segundo Pinto (2006, p.12), um sistema formal, considerado em si mesmo, constitui um conjunto de relações entre símbolos em que, neste estágio, ainda não foi estabelecida uma conexão entre o **sistema formal** e a **realidade**. Ainda não se sabe sobre quais entidades tratam as proposições do sistema. Uma vez construído um sistema formal em nível puramente sintático, pode-se estabelecer uma conexão entre ele a realidade, através de uma **interpretação**. Do ponto de vista semântico, pode-se construir uma interpretação para o sistema formal através de um modelo. Para ver como isso pode ser feito, uma linguagem formal é construída e, depois, se estabelece uma conexão entre esta linguagem e algum domínio da realidade, através de uma **semântica formal**.

A partir das breves explanações, as seções 2.1.1 e 2.1.2 apresentam contribuições de dois campos que permitiram à Ciência da Informação bem como à Ciência da Computação evoluir nos estudos sobre *representação do conhecimento*: a lógica juntamente com a filosofia clássica e a teoria da terminologia, respectivamente.

2.1.1 Contribuições da Lógica e da Filosofia Clássica

O termo *ontologia* tem sua origem na Filosofia, sendo assim, julga-se pertinente analisar a sua contextualização nesse campo do conhecimento, além de mostrar como a Filosofia clássica contribuiu e ainda contribui em seus aspectos epistemológicos para as ciências da Informação e da Computação.

Corazzon (2008) apresenta definições do termo ontologia segundo alguns filósofos renomados do campo. Segundo Albertazzi (1996)[1] apud Corazzon (2008), o termo ontologia tem sido estudado de dois modos na filosofia contemporânea: formal e material. Na primeira abordagem, a ontologia formal usa instrumentos como a lógica formal[2] para interpretar as características lógicas das asserções e as várias teorias do universo; a segunda abordagem trata a ontologia como sendo material[3], ou seja, analisa as categorias fundamentais dos objetos (ou simplesmente as coisas), seus estados, a relação entre o todo e suas partes e regras de dependência. Tal definição nos leva a uma visão de que ontologia seria um ramo da Filosofia ocupado em investigar as coisas do mundo em seus aspectos metafísicos ou abstratos.

O conceito de abstração das coisas do mundo pode ser explicado pelo filósofo Aristóteles[4], que, ao conceituar uma "coisa", separava os seus aspectos acidentais (atributos específicos das coisas – concebe o termo *differentia*) para então encontrar suas características comuns e essenciais. Esse processo intelectual através dos sentidos é considerado por Aristóteles uma *abstração* (ZILLES, 1998) de parte da realidade, na qual os objetos são analisados em seus aspectos mais gerais e classificados a partir de suas semelhanças e diferenças. No âmbito das ontologias de sistemas de informação (SMITH, 2004), conforme

[1] ALBERTAZZI, L. *Formal and material ontology*. In: POLI, R.; SIMONS, P. (Ed.). Formal Ontology. Dordrecht: Kluwer, 1996. p.199-232.

[2] "[...] using the tools and approach of formal logic: from this point of view formal ontology examines the logical features of predication and of the various theories of universals. The use of the specific paradigm of the set theory applied to predication, moreover, conditions its interpretation" .

[3] "[…] analyses the fundamental categories of object, state of affairs, part, whole, and so forth, as well as the relations between parts and the whole and their laws of dependence".

[4] Considerado o mais poderoso e influente pensador da Grécia clássica (384-322 a.C.), ao lado de seu mestre Platão. Aristóteles dedicou-se ao estudo dos seres vivos e da natureza. Fez da biologia o modelo de suas investigações (ZILES, 1998, p. 62).

será visto mais adiante neste trabalho, a abstração é uma atividade intelectual demandada na representação de conceitos e relações entre conceitos que irão compor a ontologia. Tal atividade, apesar de possuir uma abordagem geralmente pragmática, pode vir a ser desenvolvida com fundamentos epistemológicos de outras ciências como a Filosofia, que busca compreender as formas pelas quais os homens conhecem as coisas do mundo (BRAGA, et al., 2003).

A compreensão da realidade através da representação do conhecimento fez com que Aristóteles se ocupasse em analisar primeiramente a linguagem, pois considerava que a linguagem é a maneira de exprimir conhecimento do que realmente existe. Classifica, então, as expressões em categorias, dentre as quais se destacam a de: substância, quantidade, qualidade, tempo, lugar, paixão, relação, estado, ação e posição (CORAZZON, 2008a). As categorias ajudariam na organização de objetos ou coisas do mundo, uma vez que eles estariam classificados em tais categorias. Dessa forma, o raciocínio sobre os objetos poderia ser conduzido de uma maneira mais lógica e intuitiva.

Além dessas dez categorias, Aristóteles ainda classifica os vários tipos de proposições[5] em *universais* (todos os homens são mortais), *particulares* (alguns homens são bons) e *singulares* (Platão é o autor de *A República*). E ainda classifica as substâncias em gênero e espécie. O gênero apresenta propriedades em comum entre substâncias diferentes (o homem e o cavalo são animais), já a espécie apresenta diferenças entre um gênero (o homem é um animal *político*). Segundo Zilles (1998), as proposições, regidas pela evidência, formam o conhecimento através de uma regra lógica, denominada por Aristóteles de *silogismo*. Um exemplo clássico de silogismo seria: "Todo homem é mortal; Sócrates é homem; logo Sócrates é mortal". Tal exemplo mostra duas proposições consideradas verdadeiras que levam a uma terceira "logicamente" verdadeira (não necessariamente seria uma proposição verdadeira perante a realidade do mundo). Essas leis do pensamento deveriam comandar as operações da mente. Desta maneira, deu-se início ao campo chamado *lógica*, que se preocupa em expressar o conhecimento, isto é, o raciocínio humano, sobre todos os tipos de coisas no mundo e sobre as relações entre elas.

Segundo Copi (1981), o raciocínio faz parte do pensamento, no qual se realizam inferências ou se derivam conclusões a partir de proposições. Estas, por sua vez, não são peculiares a nenhuma linguagem na qual podem ser expressas. As sentenças (ou orações declarativas) na Figura 1:

---

[5] Usadas para designar o significado de uma sentença ou oração declarativa (COPI, 1981, p.22)

Chove. *It is raining.* *Il pleut.* *Es regnet.*

Figura 1 – Orações declarativas dispostas de modo diferente
Fonte: Copi (1981, p.22)

Todas têm o mesmo significado, apesar de terem sido enunciadas em português, inglês, francês e alemão, respectivamente. Isso significa dizer que, num mesmo contexto sentenças diferentes (dispostas de modo diferente) podem ter o mesmo significado e expressar a mesma proposição. Além disso, em diferentes contextos, uma única sentença pode ser usada para fazer declarações diferentes. No contexto das ontologias, tal característica possibilitaria elaborar proposições que refletissem o significado de um conjunto de modelos pretendidos em uma estrutura intensional (GUARINO, 1998), e, em contextos apropriados, tais proposições poderiam ser usadas para fazer diferentes declarações, ou seja, a depender da linguagem do modelo, conhecida por estrutura extensional.

Essas abordagens aristotélicas e da lógica vieram a contribuir mais tarde em áreas como a Inteligência Artificial, principalmente no que tange aos formalismos de representação, além do uso de categorias para organização de conceitos presentes nas bases de conhecimento (REZENDE, 2003; RUSSELL e NORVIG, 2004). Na Ciência da Informação, as categorias aristotélicas também foram aplicadas nos anos de 1930 quando Ranganathan[6], um importante estudioso da teoria da classificação facetada, concebe algumas categorias para classificar assuntos bibliográficos. Tais categorias foram embasadas nas categorias de Aristóteles e, atualmente, são aplicadas no âmbito das linguagens documentárias (DAHLBERG, 1978; CAMPOS, 2001).

2.1.2 Contribuições da Terminologia

Na concepção de Campos (2001), a Teoria da Terminologia diz respeito a uma disciplina científica que, através de princípios metodológicos, consegue elaborar terminologias – sistemas de conceitos – estruturadas para uma área de conhecimento. Tais princípios seriam: a) identificação do objeto concreto ou abstrato; b) análise das características[7] intrínsecas e extrínsecas do objeto de forma a definir o conceito e as relações

[6] A teoria da classificação facetada de Ranganathan pode ser conferida na seção 2.2.3.1.1.
[7] As características intrínsecas são as que constituem o objeto em si, tais como forma, tamanho, material, cor, etc; as características extrínsecas referem-se ao uso ou função do objeto, sua origem ou outras relações com objetos diferentes.

entre conceitos, isto é, formar sentenças corretas sobre o objeto; c) construção do sistema de conceitos a partir do princípio de contextualização ou universo de conhecimento, em que as definições dos conceitos e seus posicionamentos no sistema estão diretamente ligados à área de conhecimento na qual a terminologia está sendo construída; e d) definição dos termos como forma de expressar as unidades de pensamento ou conhecimento (conceitos) presentes no sistema.

Para a determinação de um conceito dentro de um sistema de conceitos, trabalha-se com as suas características, que segundo Dahlberg (1978), podem ser de natureza intensional e extensional. A intensão seria o somatório das características que constituem um conceito, já a extensão seria a totalidade ou o número de conceitos que um conceito atinge. Na visão de Guarino (1998), as ontologias possuem características intensional e extensional, pois são constituídas de um vocabulário consensual formando uma rede conceitual de caráter intensional, além de possuírem extensões acerca de seus conceitos, conforme poderá ser visto na seção 2.3.3.1.

Após a definição do conceito através de sua intensão e extensão, o próximo passo seria classificá-lo em um sistema de conceitos em grupos de categorias previamente concebidos por meio da atividade de categorização (RANGANATHAN 1967; DAHLBERG, 1978). Segundo Ranganathan (1967), tal agrupamento se dá quando se identifica uma característica comum aos conceitos, formando classes genéricas e específicas de conceitos, também chamadas de cadeias de conceitos, as quais organizam hierarquicamente uma estrutura classificatória em gênero-espécie e todo-parte. Além da relação lateral, ou seja, que forma renques de conceitos.

Na teoria da Terminologia, as relações no sistema de conceitos podem ser *lógicas* e *ontológicas*. As relações lógicas dividem-se em dois grupos: relações *lógicas de comparação* e relações *de combinação lógica*. Tais relações são apresentadas nas Tabelas 1 e 2.

Relações lógicas de comparação	Definição
Subordinação lógica	Um conceito tem todas as características do outro conceito, e este último possui uma característica adicional. Pode-se dizer que este conceito é uma espécie do outro, o gênero. Exemplo: funcionário (generalização) e secretária (especialização).
Coordenação lógica	Quando dois conceitos analisados são específicos do mesmo termo genérico, distinguindo-se por uma única característica. Exemplo: técnico e engenheiro, ambos seriam especializações de funcionário.
Interseção lógica	Quando são comparados dois conceitos, cujas características são idênticas parcialmente, isto é, nem todas as características são as mesmas como, por exemplo, ensino – instrução.
Diagonal lógica	Quando duas espécies de conceitos do mesmo gênero não estão ligadas por relação de subordinação nem de coordenação como, por exemplo, balão/foguete, navio/avião.

Tabela 1 – Relações lógicas de comparação
Fonte: adaptado de Campos (2001, pgs. 74-78)

As relações de *combinação lógica* se dão entre três ou mais conceitos. Podemos considerar dois dentre eles como conceitos de partida que, por sua ligação, constituem um terceiro conceito. A combinação lógica possui três espécies apresentadas na Tabela 2.

Relações de combinação lógica	Definição
Determinação	Ocorre quando um segundo conceito aparece na compreensão de um primeiro conceito como característica suplementar. Como resultado aparece um terceiro conceito, que é um específico do primeiro conceito de partida. Exemplo: a combinação dos conceitos *importar* e *homem de negócios* resulta no conceito *importador*.
Conjunção	Quando as características de dois conceitos são reunidas, resultando em um terceiro conceito que é específico comum aos dois conceitos de partida. Exemplo: a combinação dos conceitos *engenheiro* e *gerente* pode resultar no conceito *engenheiro-gerente.* Desta forma, um *engenheiro-gerente* é um engenheiro como também um gerente.
Disjunção	É definida como uma soma lógica. Somam-se os dois conceitos específicos, resultando um conceito genérico comum dos dois conceitos de partida. Exemplo: homem, mulher = ser humano adulto.

Tabela 2 – Relações de combinação lógica
Fonte: adaptado de Campos (2001, pgs. 74-78)

As relações *ontológicas* são relações indiretas entre os conceitos, pois resultam das propriedades que possuem os objetos do mundo real. Caracterizam-se pela conexão entre

causas e efeitos, ou entre o todo e suas partes. As relações, apresentadas na Tabela 3, dividem-se em dois grupos: relações de *contato* e relações de *causalidade*.

Relações ontológicas	**Definição**
Relações de contato	São relações ontológicas que se auto-explicam a partir das suas espécies, denominadas relações de coordenação ou partitiva. Exemplo: avião/motor
Relações de causalidade	Baseiam-se em um elo sucessivo de causas. As principais relações de causalidade são as relações genealógicas como pai - filho. Outros exemplos: corrosão – ferrugem (causa - efeito); petróleo – plástico (material - produto); telefone – transmissão (instrumento - função).

Tabela 3 – Relações ontológicas
Fonte: adaptado de Campos (2001, pgs. 74-78)

Campos (2001) enfatiza que o engenheiro austríaco Eugen Wuester foi o precursor dos estudos voltados à terminologia nos anos 30 ao organizar a terminologia de Eletrotécnica, tendo como objetivo facilitar a comunicação nessa área de conhecimento. A autora considera que a terminologia, por se tratar de uma linguagem artificial (configurada dentro de uma comunidade de especialistas), pode evitar a ocorrência de polissemia[8], sinonímia[9] e homonímia[10], características presentes na Lexicologia, na qual lida com a linguagem natural.

2.2 Contribuições da Ciência da Informação para a Representação da Informação

Nesta seção busca-se uma reflexão teórica acerca das principais temáticas que circundam o campo de Ciência da Informação, principalmente na linha Organização e Uso da Informação[11], na qual esta pesquisa está inserida. Os assuntos abordados procuram mostrar como a Ciência da Informação emprega aspectos teóricos e metodológicos na concepção de instrumentos de representação de informação como os vocabulários controlados. Tal característica permite a esta ciência cumprir o seu principal propósito de levar o usuário ao encontro de suas necessidades informacionais, conforme assinala Saracevic (1996, p.47): “um

[8] Quando uma palavra apresenta vários significados (ANSI, 2005).
[9] Quando um significado determina várias palavras ou termos (ANSI, 2005).
[10] Quando duas palavras apresentam a mesma forma (fonética e gráfica), mas têm dois significados diferentes (ANSI, 2005).
[11] Na Universidade Federal de Minas Gerais, tal linha encontra-se na área de concentração de produção, organização e utilização da informação (OLIVEIRA, 2005). Destacam-se os temas: Representação da informação (classificação, descrição e modelagem) em contextos digitais, análise de assunto, Bibliometria, estudos de usos e usuários de sistemas de informação.

campo dedicado às pesquisas científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos".

Contudo, percebe-se um campo interdisciplinar que possibilita outros campos como a filosofia, a terminologia, a computação, entre outros, contribuírem através de teorias, métodos e técnicas na elaboração de instrumentos cada vez mais eficientes.

A seção está organizada da seguinte forma: a seção 2.2.1 trata da principal questão do campo, a organização da informação; a seção 2.2.2 apresenta os sistemas de recuperação de informação e seus principais subsistemas; a seção 2.2.3 tece considerações importantes sobre as linguagens documentárias, destacando as teorias que as fundamentam, além de princípios metodológicos utilizados na construção de seus vocabulários controlados.

2.2.1 Organização da Informação

Como precursor da mudança na forma de organização da informação e meios para recuperá-la pode-se destacar Vannevar Bush[12] que já na década de 40 afirmava que a mente humana trabalha por meio de associações. Quando uma pessoa se depara com um objeto, logo o relaciona com um próximo através da associação de idéias (BUSH, 1945). Esse processo cognitivo deve-se ao conceito que cada pessoa tem perante uma idéia. Desta maneira, Bush propõe uma técnica na seleção da informação, que indexaria de forma associativa um objeto ao outro – a chamada máquina *Memex*. Alguns teóricos postulam que a Ciência da Informação, enquanto campo científico, surge nesta época e sob essa influência, assim como a Inteligência Artificial (SARACEVIC, 1996).

Na Ciência da Informação, a área da organização da informação tem como objetivo elaborar métodos e instrumentos para a representação de informações em diversas áreas do conhecimento. Tal representação possibilitaria, assim, a comunicação entre os usuários de um sistema de informação e o próprio sistema (NOVELINO, 1996; CAMPOS, 2001). Segundo Lima (2004), na atividade de organizar, representar e recuperar a informação, a mesma passa por processos que são altamente influenciados pela cognição humana. Tais processos, segundo a autora, seriam: a) categorização; b) indexação; c) recuperação da informação e d) interação homem-máquina. Nas seções adiante esses processos são mostrados com maior especificidade.

---

[12] Cientista do Instituto Tecnológico de Massachusetts (em inglês *Massachusetts Institute of Technology*, MIT) e chefe do esforço científico americano durante a segunda guerra mundial. Bush identificou o problema da explosão informacional, principalmente em ciência e tecnologia, e se propôs a tornar acessível um acervo crescente de conhecimento da época (SARACEVIC, 1996).

2.2.2 Sistemas de Recuperação da Informação

Os sistemas de recuperação de informação vêm se destacando na sociedade contemporânea, principalmente com o advento da Internet e da Web. O surgimento de instrumentos para viabilizar a recuperação de informação nos mais diversos ambientes, como a própria Web, bibliotecas, museus, justificam-se na medida em que buscam padronizar a entrada de dados, facilitar a estratégia de busca e, consequentemente, melhorar a interação do usuário com o sistema de recuperação de informação. Nesse sentido, algumas etapas de padronização são elaboradas no intuito de traduzir a linguagem natural de documentos[13] (considerados informação em qualquer formato), dos usuários e dos indexadores em uma linguagem de recuperação de informação.

Antes de adentrar no assunto em pauta é válido elucidar a Teoria Geral de Sistemas (TGS) como precursora das teorias sobre sistemas de informação e sistemas de recuperação de informação. A TGS surgiu com os trabalhos do biólogo Ludwig von Bertalanffy (BERTALANFFY, 1968), que na década de 30 concebeu idéias sobre teorias e formulações conceituais sobre sistemas que pudessem criar condições de aplicações em diversas áreas do conhecimento. Já na década de 50, Bertalanffy destaca que seria interessante abordar a questão do todo no sentido de identificar as partes constitutivas do sistema e os relacionamentos entre tais partes. E define que o somatório das partes e o relacionamento entre elas constituiria o todo sistêmico. Bertalanffy, então, define *sistema* como um conjunto de elementos interdependentes e interagentes que formam um todo organizado e cujo resultado seria maior do que o resultado que os elementos poderiam ter se funcionassem independentemente. Assim, a TGS se propõe como uma ciência da totalidade aplicável a todas as ciências que tratam de “todos organizados” (VASCONCELLOS, 2002).

Os SIs (Sistemas de Informação) e os SRIs (Sistemas de Recuperação de Informações) se encaixam nessa perspectiva de “todo organizado” pelo fato de possuírem componentes que trabalham juntos para cumprir um determinado propósito. O conceito de SI é mais genérico do que o de SRIs, isto é, um SRI é sempre um caso particular de um SI. Um SI é um sistema no qual a entrada principal são os dados, que através da seleção, tratamento e organização são transformados em informação para algum usuário. Araujo (1994) afirma que um SI objetiva a realização de processos de comunicação, permitindo o acesso, por seus usuários, a informações neles registradas. Já a proposta de um SRI é levar o usuário ao encontro de suas

---

[13] No âmbito de um sistema de recuperação, o termo documento é usado no seu sentido mais amplo, incluindo quaisquer outros objetos informativos (BUCKLAND, 1991).

necessidades de informação através das tarefas de: a) *representação* das informações contidas nos documentos: usualmente através dos processos de indexação e descrição dos documentos; b) *armazenament*o: gestão física e/ou lógica dos documentos e de suas representações; e c) *organização* e *acesso* aos itens de informação: recuperação das informações e dos próprios documentos armazenados, de forma a satisfazer as necessidades de informações dos usuários (SOUZA, 2005).

Um sistema de recuperação de informação, de acordo com a Figura 2, possui dois subsistemas principais: a *entrada* e a *saída*, subdivididos em mais seis subsistemas – de documentos, de indexação, de vocabulário, de busca, de interface com o usuário e de tradução.

O subsistema de *entrada* refere-se ao processo intelectual de escolha dos documentos e na descrição de seus conteúdos, que devem compor as bases de dados. A atividade de descrição de documentos é realizada pelo processo de indexação. A indexação é definida por Lancaster e Warner (1993) como um processo intelectual de analisar os assuntos ou conteúdos de um documento (análise conceitual) para sua posterior transformação (tradução) em um vocabulário ou linguagem de indexação. Em alguns sistemas, isso implica na construção e disponibilização de um vocabulário controlado como tesauros e taxonomias a fim de representar o assunto de um documento. Após o término da indexação, os documentos são arquivados em alguma base de dados de documentos e os registros contendo as representações dos documentos são direcionados a uma outra base de dados para que possam ser pesquisados. Segundo Araujo (1994), a principal fonte de realimentação (também denominado *feedback*) desse subsistema vem do subsistema de disseminação, que inclui o uso dos documentos pelos usuários.

No subsistema de *saída*, Lancaster (1993) considera-o como a interface do usuário com o sistema, partindo da necessidade de informação, passando pela estratégia de busca, de modo a propiciar a recuperação de itens relevantes[14] e, finalmente, o acesso ao documento pelo usuário. As requisições ou perguntas dos usuários passam por uma análise conceitual e são traduzidas para o vocabulário do sistema. Após essa etapa, a expressão de busca é formulada e o sistema compara as representações dos documentos com as requisições dos

[14] Os documentos mais relevantes são aqueles que possuem um conjunto de palavras mais parecido com o conjunto de palavras da necessidade de informação. Na avaliação da recuperação de um documento são considerados dois conceitos fundamentais: a revocação (do inglês *recall*), que considera a relação entre documentos relevantes recuperados e o número total de documentos relevantes existentes; e a *precisão* (do inglês, *precision*), que determina a relação entre documentos recuperados relevantes e o número total de documentos recuperados. A idéia é manter uma boa relação entre precisão e revocação, evitando assim falsos positivos e falsos negativos (ABADAL e CODINA, 2005).

usuários. Araujo (1994) afirma que a realimentação desse subsistema provém do subsistema de estratégia de busca, que pode refletir na decisão do usuário em modificar a sua estratégia de busca com base nos documentos recuperados. No caso do subsistema de avaliação, o usuário julga a relevância do documento recuperado.

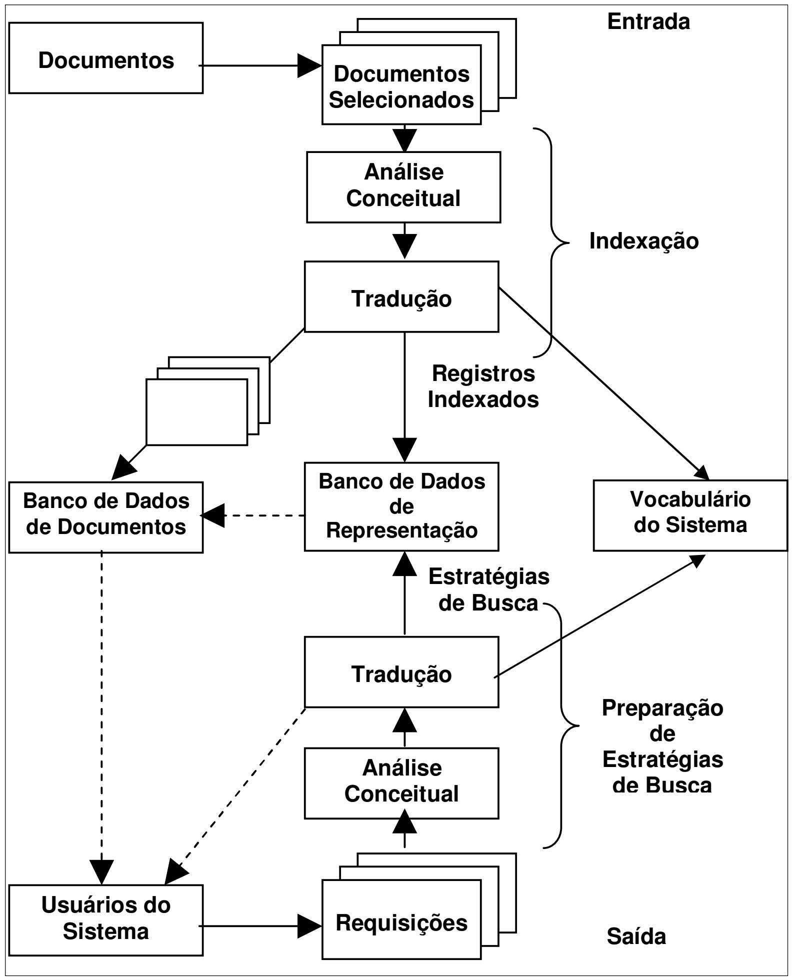


Figura 2 - Sistema de Recuperação de Informação
Fonte: Lancaster (1993)

Numa abordagem contemporânea, Abadal e Codina (2005) explicitam sobre as operações envolvidas num processo de recuperação de informação que podem vir a ser humanas, automatizadas[15] ou uma combinação das duas. Os autores propõem ainda um modelo universal para tal processo: a) seleção dos documentos: informação de relevância para o público de interesse; b) indexação: escolha dos conceitos relevantes do domínio, sendo considerado pelos autores um trabalho de natureza intelectual; c) ordem de apresentação dos documentos ao usuário (também chamado de *ranking*); d) estabelecimento da estrutura de

[15] O processo automático consiste na execução pelo computador de algoritmos de recuperação de informação. Maiores informações sobre o assunto podem ser encontradas em Souza (2006).

navegação entre os documentos (utilização da tecnologia hipertexto); e) definição de uma taxonomia para os documentos (também chamado de categorização); f) produção de resumos para os documentos; e g) representação em formato gráfico das informações contidas nos documentos. O modelo pode ser visualizado na Figura 3, na qual representa duas etapas: o lado do usuário e o lado do sistema.

NI -> *Análise* -> RNI -> *Comparação* <- RTEXTO <- *Análise* <- TEXTO

NI: necessidade de informação (a pergunta do usuário)
Análise: totalmente automático, totalmente intelectual, ou uma combinação dos dois
RNI: representação da necessidade de informação (representação da pergunta)
Comparação: função que gera os resultados mediante a comparação da representação da pergunta com a representação do texto
RTexto: representação do texto (representação do documento)
TEXTO: documento

Figura 3 – Modelo Universal de um SRI
Fonte: adaptado de Abadal e Codina (2005)

No que tange ao lado do usuário, a figura mostra a necessidade de informação do mesmo (*NI*); a análise conceitual (*análise*) que pode ser automática, intelectual ou uma combinação das duas; e a representação da pergunta feita pelo usuário (*RNI*); já o lado do sistema tem-se o documento (*texto*); a análise conceitual e representação do documento (*Rtexto*); e a comparação, considerada o núcleo do sistema, isto é, a parte principal do processo de recuperação, na qual irá comparar a representação da pergunta do usuário com a representação da base documental disponível. Tal processo consiste em verificar os elementos homogêneos existentes entre os dois lados – usuário e sistema – e medir o grau de relevância existente na recuperação.

Uma ontologia busca padronizar a linguagem envolvida no domínio e facilitar a comunicação entre diversos usuários com visões diferentes do processo, através de um vocabulário compartilhado do domínio de interesse (GUARINO, 1998; USCHOLD e GRUNINGER, 1996). Para tal, as ontologias poderiam auxiliar o trabalho do indexador no momento da representação[16] dos documentos, pois são sistemas constituídos por um vocabulário e um conjunto de definições consensuais no âmbito de um domínio, que pode se apresentar em uma linguagem lógica. Esta, por sua vez, através de sua semântica formal, poderia eliminar contradições envolvendo os conceitos e as relações entre os conceitos, resultando numa especificação não ambígua do domínio. E, finalmente, com base no

[16] Uma representação pode cobrir aspectos intrínsecos (conteúdo) e extrínsecos (catálogo) do documento.

formalismo, as pessoas envolvidas num SRI poderiam se comunicar sob o mais alto nível de abstração a respeito de um determinado domínio, viabilizando, em função disso, a estratégia de busca e a recuperação da informação por parte dos usuários. As ontologias funcionariam, desta forma, como uma interface entre a necessidade do usuário e a representação de documentos.

Assim como as ontologias, outros instrumentos são usados na organização de informação de SRIs. Estes instrumentos, em geral, denominados linguagens documentárias, têm a proposta de levar o usuário ao encontro de suas necessidades de informação. Na próxima seção são apresentadas as principais teorias e princípios metodológicos da Ciência da Informação que abarcam a construção de modelos conceituais, que permitem a elaboração de linguagens documentárias.

2.2.3 Linguagens Documentárias e Teorias Subjacentes

As linguagens documentárias compreendem a comunicação entre usuários e sistemas de informação. São linguagens artificialmente construídas a partir da linguagem natural presente nos documentos, buscando-se obter um vocabulário controlado de um assunto específico (DODEBEI, 2002). Podemos citar alguns desses vocabulários: os esquemas de classificação, os tesauros e as taxonomias. Para tal, teorias ligadas à representação de conceitos, como a teoria da classificação facetada e a teoria do conceito, tornam-se possibilidades teórico-metodológicas para construção de linguagens documentárias. A teoria da classificação é apresentada na seção 2.2.3.1, a teoria do conceito é elucidada na seção 2.2.3.2 e a seção 2.2.3.3 apresenta os vocabulários controlados: esquemas de classificação bibliográfica, listas de cabeçalhos de assunto, tesauros e taxonomias.

2.2.3.1 Teoria da Classificação

Dahlberg (1976) destaca como o pai da moderna Teoria da Classificação o matemático e bibliotecário indiano Shialy Ramamrita Ranganathan, que em 1933 concebeu a Teoria da Classificação Facetada, a qual teve sua inspiração nos trabalhos do *classificacionista* Henry Evelyn Bliss dos Estados Unidos. O sistema de Ranganathan se diferenciou dos outros esquemas de classificação (apresentados brevemente na seção 2.2.3.3.1) pelo fato de não

trabalhar com classes pré-estabelecidas, advindas da teoria descritiva[17], visto que estas precisavam adequar os assuntos tratados nos documentos à estrutura classificatória existente nos esquemas. No esquema facetado, a classificação dos livros era criada somente no momento em que um livro fosse analisado de acordo com os elementos constitutivos de seu assunto. O conhecimento para Ranganathan era visto como algo dinâmico, sendo necessário desenvolver uma teoria que fosse capaz de superar algumas barreiras apresentadas nos esquemas de classificação até então em uso. Surgia, assim, a *Teoria Dinâmica do Conhecimento*. Tal teoria justifica-se em ser apresentada em função de ser objeto de investigação da presente pesquisa, visto que, no processo de construção de ontologias, a análise facetada pode ser empregada na fase de modelagem do domínio. Segundo Prieto-Díaz (2003, p.5), "um esquema de classificação [...] pode ser considerado uma taxonomia estendida ou uma ontologia reduzida" e ainda " [...] técnicas derivadas de esquemas de classificação podem ser usadas para sistematicamente iniciar a criação de ontologias" [18].

Segundo Campos (2001), a *Teoria Dinâmica do Conhecimento* propõe organizar assuntos novos e já conhecidos em lugares apropriados no esquema de classificação. Tal teoria encontra-se explicitada através dos princípios, cânones e postulados nos *Prolegomena to library classification* (RANGANATHAN, 1967) publicado por Ranganathan pela primeira vez em 1933, a Classificação de Dois Pontos, conhecida como *Colon Classification*[19]; em seguida, uma segunda edição em 1957, descrevendo a Teoria Dinâmica da Classificação Bibliográfica; e uma terceira edição em 1967, contendo uma versão mais avançada da Teoria Dinâmica.

Os princípios da Teoria Dinâmica influenciaram um novo tipo de Classificação Bibliográfica, a Classificação Facetada. A Classificação Facetada divide um assunto por seus múltiplos aspectos ou facetas, isto é, em grupos de classes[20] reunidas por um mesmo princípio de divisão (BARBOSA, 1972; DAHLBERG, 1976; VICKERY, 1980, GUINCHAT e MENOU, 1994).

---

[17] Descreve o estado atual do conhecimento e não possui mecanismos que permitam atender as mudanças advindas das diversas áreas do conhecimento (CAMPOS, 2001).

[18] "A classification scheme […] can be considered an extended taxonomy or a reduced ontology." "[…] techniques for deriving classification schemes can be used for systematically initiating the creation of ontologies."

[19] Tabela de classificação elaborada para a organização do acervo da Biblioteca da Universidade de Madras, na Índia (CAMPOS, 2001, p.27)

[20] Barbosa (1972) define classe como um conjunto de coisas que apresentam algo em comum e, quando aplicada uma diferença, resulta numa subclasse.

Dahlberg (1976) ressalta que para classificar um documento faz-se necessária uma análise temática ou um tratamento descritivo sobre seu conteúdo, ou seja, uma análise conceitual do documento. Após a análise, é feita a ordenação desses elementos (denominados isolados) em facetas, as quais representam uma coleção de termos com relacionamento hierarquicamente igual ao assunto geral, com princípios de divisões comuns. Cada elemento pertencente à faceta é denominado um foco. Tal abordagem é denominada por Dahlberg (1976) de método analítico-sintético.

Um conjunto de facetas com características comuns é chamado de categoria, que é uma divisão altamente generalizada do conhecimento. Barite (2000, pgs. 5 e 6) explicita que "[...] categorias são relevantes como instrumentos de análise e organização de objetos, fenômeno e conhecimento". E ainda que "[...] a noção de categoria facilita o processo de análise de assunto [...] ajuda no estabelecimento da correta precedência entre os vários assuntos nos documentos [...]" [21]. Jacob (2004, p.518) afirma que o mundo é dividido em grupos ou categorias que compartilham alguma similaridade dentro de um domínio. As categorias especificadas podem variar de domínio para domínio de acordo com o poder de abstração de quem está modelando o sistema. Desta maneira, os aspectos cognitivos tornam-se importantes nesse processo.

Campos (2001) afirma que a premissa básica para entender a Classificação Facetada seria compreender a definição de idéia, informação, conhecimento e assunto. A autora descreve tal definição segundo a visão de Ranganathan, no qual diz ser a idéia um produto do pensamento, da reflexão, da imaginação, que, com a ajuda da lógica do intelecto, integra uma seleção de percepções. A informação se constituiria no momento em que a idéia é comunicada a alguém ou obtida por meio do estudo pessoal e da investigação. O conhecimento seria o conjunto de todas as idéias conservadas pela humanidade. E, por fim, o assunto é visto como um conjunto de idéias organizadas e sistematizadas num campo de interesse específico.

Diante desse conjunto de idéias organizadas e sistematizadas, Ranganathan cria cinco idéias fundamentais para serem utilizadas na análise dos assuntos contidos nos documentos, de forma a organizar os componentes desses assuntos em um sistema de classificação - as chamadas categorias fundamentais para a Classificação Facetada, conhecidas pelo acrônimo PMEST – Personalidade; Matéria; Energia; Espaço; e Tempo. A categoria Tempo é

---

[21] "[…] categories are only relevant as instruments of analysis and organization of objects, phenomena and knowledge". E ainda que "[…] the notion of category facilitates the subject analysis process […] helps to establish correct precedence among several subjects in a documents […]"

considerada a mais fácil em termos de visualização, por ser evidente em si mesma (período associado com o assunto). A categoria Espaço é geralmente uma área geográfica da localização de um assunto. Já a categoria Energia pode ser reconhecida pela conotação de ação. A categoria Matéria manifesta em si mesma como material e propriedade. E, por fim, a categoria Personalidade, considerada por Ranganathan como indefinível. Tal categoria é considerada a mais difícil de ser identificada; geralmente reconhecida por critérios de eliminação, ou seja, após a identificação das categorias Espaço, Tempo, Matéria e Energia em um determinado assunto, a faceta será enquadrada na categoria Personalidade. Esse método é chamado de Método de Resíduos (RANGANATHAN, 1985).

Além das categorias, Ranganathan introduz três níveis distintos no processo de criação de sistemas de classificação: da Idéia, o Verbal e o Notacional. O plano de idéias é considerado um trabalho de análise do assunto, considerado um processo pelo qual se estabelecem as relações a serem reconhecidas entre os conceitos; a definição da sintaxe, ou a relação entre os termos da linguagem, seria o plano notacional. E a padronização da terminologia para servir de comunicação das idéias ou conceitos (livre de homonímia e sinonímia) pertenceria ao plano verbal.

Podemos visualizar na Teoria da Classificação Facetada a importância dada aos conceitos, que podem ser considerados como unidades de conhecimento. Tais unidades quando arranjadas de uma maneira sistematizada, através das facetas ligadas às categorias fundamentais, permitem formar qualquer assunto. O ambiente da documentação, desta forma, transcende ao Universo de Conhecimento.

2.2.3.2 Teoria do Conceito

A teoria do conceito, cunhada pela pesquisadora alemã Ingetraut Dahlberg, define “conceito” como unidade de conhecimento e que sua formação se dá pela reunião e compilação de enunciados verdadeiros a respeito de determinado objeto (DAHLBERG, 1978). Tal definição, segundo Dahlberg, possibilita um entendimento mais objetivo daquilo que está sendo observado pelo indivíduo, o qual tem a capacidade de fazer afirmações sempre verdadeiras perante as coisas reais do mundo e comunicá-las através de uma forma verbal. A visão de Dahlberg vai de encontro a um dos componentes de uma ontologia capaz de modelar sentenças verdadeiras sobre uma certa realidade, os axiomas. Tal componente permite especificar ontologias rigorosamente formais contendo “termos definidos com semântica

formal, teoremas e provas"[22] (USCHOLD e GRUNINGER, 1996, p.6). Os axiomas permitiriam, assim, definições precisas nos conceitos e relações entre conceitos, de forma a possibilitar inferências sobre o sistema de conceitos.

Alvarenga (2001) explana "conceitos" como sendo unidades de conhecimento, compostas de declarações verdadeiras sobre um item específico de referência representado por uma forma verbal, o termo. Alvarenga (2003) ainda expõe comentários sobre o trabalho de Jesse Shera, um dos pensadores clássicos da biblioteconomia, o qual trata da gênese dos conceitos na mente humana, ou seja, como se desenvolve, no cérebro humano, o processo de formação de conceitos. Esse trabalho torna-se essencial para pesquisas voltadas à Inteligência Artificial na qual se exige um profundo conhecimento do sistema nervoso humano a fim de tentar viabilizá-lo na máquina. Ainda na concepção da autora, documentos são registros do conhecimento humano que são formados de conceitos, e que o homem sempre classificou o conhecimento em uma hierarquia taxonômica que une semelhanças ou separa diferenças existentes dos conceitos. Daí a razão pela qual, segundo o referido artigo, os conceitos seriam o material para a construção da classificação.

Dahlberg (1978) apresenta três passos, mostrados na Figura 4, para a formação de conceitos:

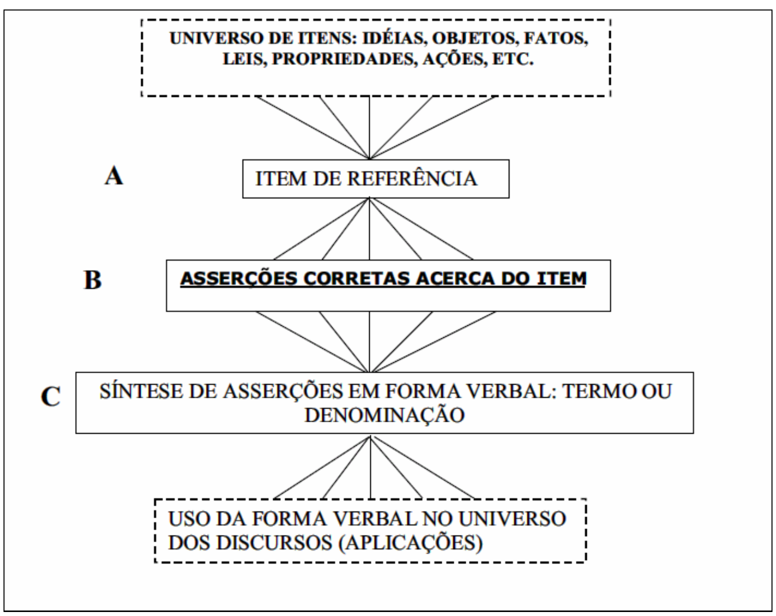


Figura 4 – Modelo para a construção dos Conceitos
Fonte: Dahlberg (1978, p.143)

i) o passo referencial, que seria o ponto de partida para o processo de determinação do conceito, ou seja, a seleção do referente num determinado domínio de conhecimento;

[22] rigorously formal […] defined terms with formal semantics, theorems and proofs"

ii) o passo predicacional, no qual há uma atribuição de características relevantes ao referente;

iii) o passo representacional, em que através das características atribuídas ao referente, consegue-se chegar à denotação do conceito, representado pelo termo.

Após, ou no momento[23] da formação ou definição do conceito, o próximo passo seria sistematizar tal unidade de conhecimento, ou seja, posicioná-la em um sistema de conceitos, utilizando-se das técnicas de categorização. Dahlberg, como Ranganathan, utiliza a noção de categoria para representar contextos, e posteriormente analisar os conceitos de um contexto no intuito de classificá-los no interior dessas categorias. A Figura 5 mostra as categorias e as subcategorias formais propostas por Dahlberg, as quais foram embasadas nas clássicas categorias de Aristóteles, que por sua vez fundamentaram as categorias de Ranganathan.

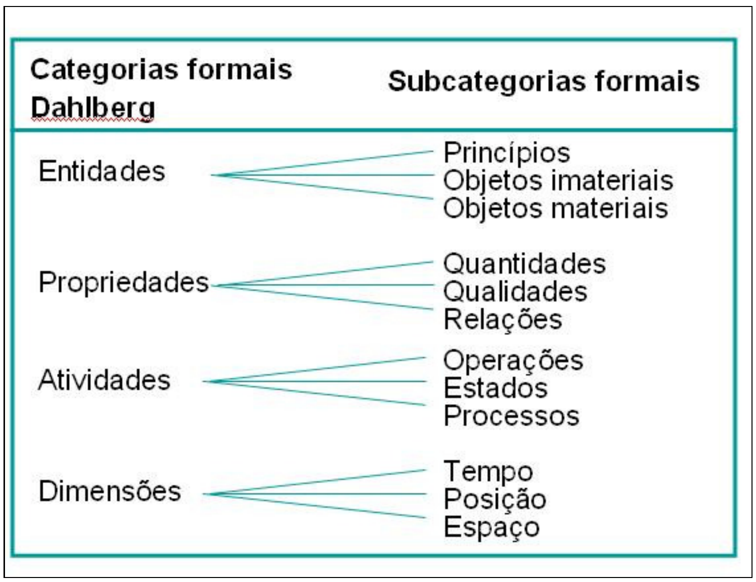


Figura 5 – Categorias e subcategorias formais de Dahlberg
Fonte: adaptado de Dahlberg (1978, p.145)

Campos (2001a, p.95) observa a técnica de categorização como: a) um recurso para o entendimento da natureza do conceito e b) um recurso de formação de estruturas conceituais. Os dois aspectos são considerados complementares e fornecem recursos para estruturar um determinado domínio de conhecimento. Diante disso, podemos conferir a relação da teoria da classificação facetada com a teoria do conceito com seu método analítico-sintético.

[23] "Pode-se afirmar que o método de categorização é o exercício de entendimento do conteúdo conceitual do termo [...] esta afirmativa nos leva a uma pergunta clássica: Define-se para classificar (categorizar) ou classifica-se (categoriza-se) para definir? Fica a questão" (CAMPOS e GOMES, 2006).

No que diz respeito às relações entre conceitos, a teoria de Dahlberg (1978) propõe as seguintes:

- Relações lógicas que podem ser categorizadas em:
  - Identidade: quando as características são as mesmas;
  - Implicação: quando os conceitos do objeto A estão contidos no objeto B;
  - Interseção: os dois conceitos coincidem em algum elemento;
  - Disjunção: os conceitos se excluem mutuamente, sem nenhuma característica em comum;
  - Negação: o conceito do objeto A inclui uma característica cuja negação se encontra no objeto B.
- Relações hierárquicas. Também conhecidas como relações gênero-espécie, subdivididas em hierárquica subordinada (cadeia) e hierárquica coordenada (renque). Quando dois conceitos possuem características idênticas e um deles possui uma característica a mais do que o outro, então é possível estabelecer a relação hierárquica entre os conceitos. Por exemplo, temos árvore, árvore frutífera, macieira.
- Relações partitivas. É quando existe uma relação entre o todo e suas partes como, por exemplo: árvore e (raízes, tronco, folhas, etc.)
- Relações de oposição. É a noção oposta entre conceitos, tal como presente e ausente, caracterizando uma contradição, e a contrariedade, como branco e preto.
- Relações funcionais. Estas relações se aplicam a conceitos que expressam processos, através do caráter semântico e suas valências do verbo. Por exemplo: produção – produto – produtor.

Campos (2001a) ressalta que os princípios da teoria do conceito estão sendo úteis para a elaboração de tesauros, por que fornecem recursos para o estabelecimento de relações, para sua realização no plano verbal e para a determinação dos termos.

Campos (2004) ainda destaca que o estudo da teoria do conceito pode ajudar o *modelizador* em sua atividade de elaboração de modelos de representação para linguagens documentárias, sistemas computacionais, hipertextos e ontologias. Tais instrumentos de representação podem proporcionar de maneira eficaz o acesso, a disseminação, o compartilhamento e uso do conhecimento pelos usuários em potencial.

2.2.3.3 Vocabulários Controlados

Vocabulário controlado é definido por Lancaster (1993) como uma lista de termos autorizados, em que o indexador somente pode atribuir a um documento termos que existem na lista adotada pela unidade de informação em que trabalha. O vocabulário controlado tem por funções: a) o controle de sinônimos, através da definição de um termo padrão, com remissivas para os sinônimos; b) a diferenciação entre os homógrafos[24]; e c) o agrupamento de termos em que os significados apresentem uma relação mais estreita entre si, como por exemplo, relações hierárquicas ou não hierárquicas. Lancaster (1993) identifica três tipos principais de vocabulário controlado: os esquemas de classificação bibliográfica, as listas de cabeçalhos de assuntos e os tesauros.

A ANSI (2005) enfatiza que o principal propósito dos vocabulários controlados é fornecer um significado para a organização da informação, incluindo tradução, consistência, relacionamentos, visualização e recuperação perante a informação.

Nesta seção são conceituados alguns instrumentos que permitem a concepção de vocabulários controlados categorizados que incluem uma análise semântica das palavras. A semântica envolvida nas linguagens documentárias busca tornar compreensivo um domínio de conhecimento através da análise conceitual do mesmo. Dentro do escopo da Biblioteconomia e da Ciência da Informação aparecem os esquemas de classificação, as listas de cabeçalhos de assunto, os tesauros e as taxonomias, instrumentos capazes de subsidiar a representação de conteúdos dos mais diversos tipos, facilitando a busca e a recuperação de informação.

De acordo com a visão de Gilchrist (2003) e Jimenez (2004), as ontologias possuem os mesmos princípios dos vocabulários controlados: trabalham com linguagem natural e fazem a delimitação de termos e de relações. Entretanto, a semântica envolvida na terminologia da ontologia se difere dos vocabulários controlados (dentro do contexto das linguagens documentárias) por incluir axiomas formais (através de declarações lógicas) que restringem a utilização do vocabulário. Nesse sentido, pode-se observar a relação das ontologias com a Ciência da Informação.

Essa seção está subdividida da seguinte forma: a seção 2.2.3.3.1 apresenta os esquemas de classificação bibliográfica; a seção 2.2.3.3.2 apresenta as listas de cabeçalho de assunto; a seção 2.2.3.3.3 apresenta os tesauros em seu aspecto histórico e atual; e finalmente a seção 2.2.3.3.4 apresenta as taxonomias.

---

[24] Vocabulário que tem a mesma grafia de outro, porém sentido diferente (ANSI, 2005).

2.2.3.3.1 Esquemas de Classificação Bibliográfica

Segundo Satija (2000), a classificação bibliográfica seria direcionada para conhecimento contido em livros e outros materiais de leitura. E teria como objetivo organizar materiais informacionais em bibliotecas ou centros de informação de modo a ser útil na consulta, localização e organização de conteúdo.

Os esquemas mais conhecidos até o início do século XX[25] eram a *Library of Congress Classification*, a *Rider's International Classification*[26], a Classificação Decimal de Dewey[27], ou como é mais conhecida, a CDD, e a Classificação Decimal Universal[28], conhecida como CDU. Tais esquemas foram concebidos a partir da teoria descritiva, na qual propunha descrever o estado do conhecimento do século XIX, sem, portanto, possuir mecanismos que permitissem atender às mudanças advindas das diversas áreas de conhecimento (CAMPOS, 2001).

2.2.3.3.2 Listas de Cabeçalhos de Assunto

Os cabeçalhos de assunto tiveram sua origem nos Estados Unidos na segunda metade do século XIX, tendo seus princípios determinados pela tecnologia vigente, que eram as fichas ou catálogos impressos (CAMPOS et al., 2002). Conforme já explanado na seção de esquemas de classificação, para organizar os livros nas estantes, as bibliotecas utilizavam a CDD, a CDU, dentre outros. E para organizar as informações dos documentos, tinham os cabeçalhos de assunto, que segundo Gomes (1996) "são um produto tipicamente americano para tratamento dos assuntos dos livros e para organização do catálogo alfabético de assunto".

Os cabeçalhos de assunto são formados de uma palavra ou de um conjunto de palavras que representam o assunto do documento. Seus elementos são considerados pré-coordenados, isto é, os termos que o compõem se coordenam em um processo prévio à sua utilização. Segundo Campos et al. (2002), o cabeçalho de assunto tem sua base na linguagem natural, ou seja, os elementos que o integram adquirem significado quando formam um contexto, sendo

---

[25] Os esquemas – tabelas com funções de enumerar assuntos do passado, do presente e do futuro antecipado - foram classificados por Ranganathan em dois períodos: o período pré-faceta, que data de 1876 a 1896, no qual os esquemas enumerativos e quase enumerativos estão inseridos; e o período de transição para a faceta, que data de 1897 a 1932, no qual os esquemas quase facetados podem ser incluídos (CAMPOS, 2001).

[26] Ambos considerados por Ranganathan e seus seguidores como esquemas enumerativos.

[27] Do termo inglês *Dewey Decimal Classification.* Considerada por Ranganathan uma classificação quase enumerativa.

[28] Do termo inglês *Universal Decimal Classification*. Considerada por Ranganathan uma classificação quase facetada.

esta uma característica impeditiva de uma representação eficaz que possa atender às necessidades do usuário em busca de informação. Segundo as autoras, diante a essa dificuldade, surgiu a necessidade de romper com o tradicional cabeçalho de assunto. A unidade de representação temática de um documento deixa de ser o assunto e passa a ser a reunião de conceitos em um sistema pós-coordenado, cujos termos são combinados no momento de seu uso. Surge, assim, um tipo de instrumento de termos controlados em sua forma e conteúdo, os tesauros.

2.2.3.3.3 Tesauros: origem, definições e vertentes de construção

O termo <u>tesauro</u> teve sua origem no dicionário de Peter Mark Roget em 1852 – *Thesaurus of English Words and Phrases*. Em tal dicionário, os termos não estão organizados em ordem alfabética, como nos dicionários tradicionais, mas de acordo com o significado das palavras que o compõem (DODEBEI, 2002). Surge, dessa maneira, uma primeira concepção de estrutura organizada logicamente através de elementos semânticos relacionados entre si. A partir de 1940, o tesauro documentário começou a ser usado na Ciência da Informação com o propósito de recuperar informação em ambientes com uma diversidade de documentos especializados.

O programa Unisist (UNESCO, 1973, p.6) define tesauro como um "vocabulário controlado dinâmico de termos relacionados semântica e genericamente, cobrindo um domínio específico do conhecimento" e ainda como "um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema mais restrita" Uma outra definição pode ser encontrada no *National Information Standards Organization* (ANSI/NISO Z39-19-1993), que define tesauro como um "vocabulário controlado organizado em uma ordem conhecida na qual as relações de **equivalência**, **hierárquicas** e **associativas** entre os termos são claramente mostradas e identificadas através de indicadores de relação padrão".

As ilustrações que se seguem mostram os tipos de relações que os tesauros comportam.

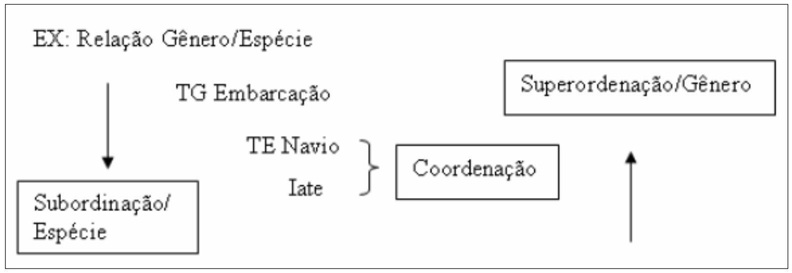


Figura 6 – Relação Gênero/Espécie
Fonte: Cintra et al. (2002)

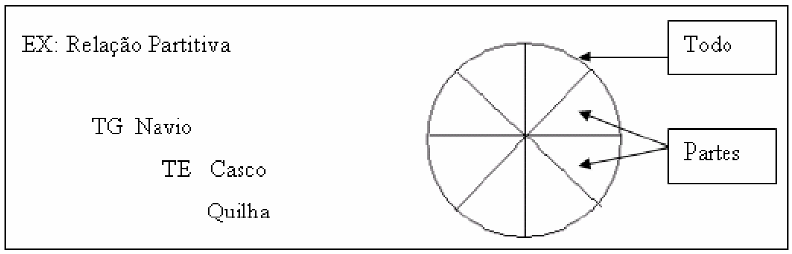


Figura 7 – Relação Todo/Parte
Fonte: Cintra et al. (2002)

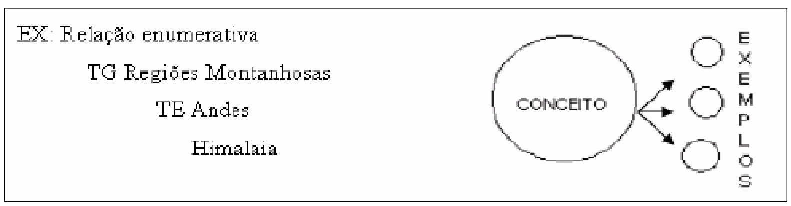


Figura 8 – Relação Enumerativa
Fonte: Cintra et al. (2002)

As Figuras 6, 7 e 8 mostram relações hierárquicas. No primeiro caso, há uma relação de subordinação entre gênero (denominado **TG** – termo geral) e espécie (denominado **TE** – termo específico). No segundo caso, há uma relação entre o todo e suas partes, conforme é mostrado na Figura 7: o termo *Casco* é uma parte e não um tipo de *Navio*. Já no terceiro caso, os termos específicos (*Andes* e *Himalaia*) são exemplos do conceito R*egiões Montanhosas*.

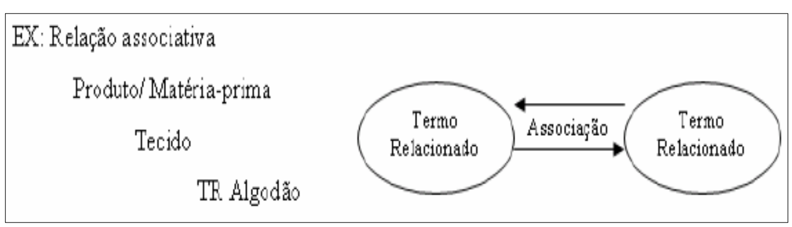


Figura 9 – Relação Associativa
Fonte: Cintra et al. (2002)

A Figura 9 mostra o tipo de relação associativa, na qual um termo pode estar relacionado a outro termo por alguma característica comum entre eles. A identificação **TR** é usada para

representar que o termo *Algodão* é um termo relacionado ao termo *Tecido* e vice-versa. As relações associativas são não-hierárquicas e teoricamente infinitas.

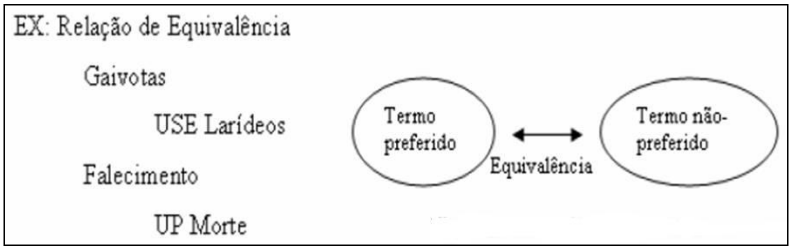


Figura 10 – Relação de Equivalência
Fonte: Cintra et al. (2002)

A Figura 10 mostra a relação de equivalência, na qual sinônimos e quase-sinônimos são remetidos aos termos envolvidos. No exemplo, a identificação **USE** é utilizada para representar que o termo *Larídeos* é sinônimo do termo *Gaivotas*, contudo Larídeos é o termo preferido. Já na identificação **UP** (usado para) o termo *Falecimento* é usado preferencialmente ao termo *Morte*.

Segundo Currás (1995), tesauro é definido como uma linguagem usada para fins documentários, composto de termos, simples ou compostos, os quais estão relacionados entre si sintática e semanticamente. A autora ressalta que os tesauros representam os assuntos dos documentos e podem auxiliar nas buscas por unidades de informação pela determinação de quais termos poderão ser usados no momento da busca; além de aproximar a linguagem do usuário à do sistema, permitindo uma introdução de novos termos e também uma alteração em sua estrutura quando se fizer necessário. Tal representação de assunto é feita pelo profissional indexador, o qual analisa o assunto do documento e depois faz o levantamento dos termos permitidos de um tesauro. Fujita (2003) destaca a análise de assunto como uma fase inicial do processo de indexação, que resultará em tópicos do assunto de um documento e posteriormente numa tradução desses tópicos em uma linguagem adotada pelo sistema, conforme foi visto na seção 2.2.2.

Em Moreira, Alvarenga e Oliveira (2004) pode ser visualizada a evolução histórica de construção de tesauros: a norte-americana e a européia, na qual as autoras embasam-se nas considerações de Lancaster (1986) para fazer tal explicitação. As autoras inserem o Unitermo (instrumento de representação do assunto por palavras únicas) como a primeira vertente, originária na América do Norte em 1951, tendo influenciado o aparecimento do primeiro tesauro. E a outra, a Teoria da Classificação Facetada, na qual continuou a ser desenvolvida pelo grupo de pesquisa denominado *Classification Research Group* (CRG) na Inglaterra. Este

grupo, formado por quatorze componentes[29], ampliou as categorias do PMEST e desenvolveu diversas tabelas de classificação na década de 50. Deu-se, então, origem a uma técnica mais refinada para construção de tesauros – denominado *Thesaurofacet*, a qual permite o melhor posicionamento do conceito dentro do sistema de conceitos em uma área de assunto específica, através do uso de categorias. As autoras mostram também uma evolução recente, a dos *tesauros-com-base-em-conceitos*, também conhecido como tesauros terminológicos, cuja origem vem da junção da teoria do conceito de Dahlberg com a Teoria da Classificação Facetada de Ranganathan. Segundo Campos (2001), os tesauros terminológicos são elaborados dando ênfase na Teoria da Classificação, Teoria do Conceito e princípios terminológicos, tendo nas características do conceito um elemento essencial para evidenciar as relações entre os conceitos e seu posicionamento no sistema, além de defini-lo, permitindo uma comunicação mais precisa entre usuário e sistema.

2.2.3.3.4 Taxonomias

Segundo a norma ANSI (2005), as taxonomias são coleções de termos classificados em uma estrutura hierárquica, na qual emprega relacionamentos de generalização e especialização.

O termo taxonomia vem sendo utilizado para designar aplicações que organizam informação em ambientes eletrônicos e virtuais (Portais Corporativos, Bibliotecas Digitais, *Intranets*, *Extranets*, etc), tendo como principais atribuições: a) auxiliar os usuários no momento da escolha dos filtros destinados à busca em grandes bancos de dados; b) ajudar os usuários na busca da informação desejada e como usar a informação recuperada; c) ajudar os usuários na navegação em sistemas hipermídia; d) possibilitar uma terminologia padrão no âmbito organizacional; e) possibilitar e facilitar o compartilhamento de informação e conhecimento nas organizações (GILCHRIST, 2003).

Vickery (1980) complementa dizendo que uma taxonomia é um sistema de classificação de itens de um domínio apresentado hierarquicamente, cuja estrutura se organiza por meio de relações genéricas, tendo como prioridade a identificação de entidades.

Algumas semelhanças podem ser percebidas entre os tesauros, as taxonomias e as ontologias, sendo todos: instrumentos utilizados para organização de informação e de conhecimento; sistemas de conceitos; sistemas com base classificatória. As taxonomias

[29] D.J.*Campell*, E.J.*Coates*, J.E.L. *Farradane*, D.J.*Foskett*, G.*Jones*, J.*Milles*, T.S. *Morgan*, B.I.*Palmer*, O.W.*Pendleton*, L.G.M.*Roberts*, B.C. *Vickery*, A.J.*Walford*, K.E.*Watkins* e A.J. *Wells* (LIMA, 2004).

identificam e classificam entidades, admitindo somente relacionamentos gênero/espécie e todo/parte, além de permitirem agregar dados em sistemas automatizados. Os tesauros representam conceitos e admitem um maior número de relações: hierárquicas, associativas e de equivalência. Já as ontologias necessitam de relações mais explicitadas, pois são instrumentos passíveis de interpretação por máquinas, diferentemente dos tesauros, concebidos, a priori, para serem manipulados por humanos.

2.3 Contribuições da Ciência da Computação para a Representação do Conhecimento

Os mecanismos de representação de conhecimento no âmbito da Ciência da Computação servem para subsidiar o processo de implementação de estruturas computacionais, que se resumem em entidades, objetos e conceitos que existem em um domínio de interesse e os relacionamentos semânticos entre tais estruturas. Nesse sentido, essa seção objetiva apresentar as principais contribuições da Ciência da Computação para a área de Representação do Conhecimento e encontra-se organizada como segue: a seção 2.3.1 apresenta a área de Inteligência Artificial elucidando seus principais formalismos de representação de conhecimento, destacando a lógica com sua notação formal capaz de desenvolver regras de inferências úteis ao raciocínio inteligente e conseqüentemente à resolução de problemas específicos; a seção 2.3.2 apresenta modelos conceituais associados à modelagem de sistemas: o modelo relacional para bancos de dados e o modelo orientado a objetos com sua linguagem de modelagem unificada, que traz contribuições no espaço comunicacional entre usuários e o modelo representado graficamente. Tais modelos são apresentados no sentido de se estabelecer uma conexão com as ontologias; e, finalmente, a seção 2.3.3 apresenta as ontologias como um instrumento capaz de definir e organizar formalmente as "coisas" que existem em um determinado domínio através de axiomas advindos de linguagens específicas para representação de ontologias (destacando as linguagens para a Web Semântica, sua aplicação mais recente). A seção aborda também os processos envolvidos na construção de ontologias, enfatizando contribuições de determinadas áreas da computação que se concentram na atividade de desenvolvimento de sistemas.

2.3.1 Inteligência Artificial

A Inteligência Artificial (IA) pode ser definida como uma área ocupada em desenvolver sistemas que procuram pensar e agir de acordo com modelos que simulem o funcionamento da mente humana (RUSSEL e NORVIG, 2004). Ainda segundo os autores, a IA juntamente com outras ciências (filosofia, matemática, ciência da computação, neurociência, lingüística, ciência cognitiva, dentre outras) procuram congregar esforços para a compreensão das complexas estruturas que compõem a mente. Nesse sentido, Russel e Norvig (2004, pgs. 4 e 5) determinam que os sistemas artificiais precisariam ter as seguintes

capacidades para uma aproximação com a mente humana : a) processamento de linguagem natural: para uma comunicação precisa em um determinado idioma; b) representação de conhecimento: para o armazenamento do que sabe ou ouve; c) raciocínio automatizado: através do conhecimento armazenado o sistema seria capaz de responder a perguntas e tirar conclusões a partir de tal conhecimento; d) aprendizado de máquina: adaptação de novas situações a partir de padrões; e) visão de computador: no sentido de perceber os objetos; e f) robótica: para movimentar objetos.

Nessa pesquisa, o interesse encontra-se na capacidade de representação de *conhecimento* desses sistemas. Chandrasekaran, Johnson e Benjamins (1999, p.23) afirmam que "na IA *conhecimento* em um sistema computacional é conceitualmente alguma coisa que é explicitamente representada e executada por um processo de inferência" [30]. Nesse sentido, a área de Engenharia do Conhecimento é elucidada como sendo uma área interessada em projetar e construir sistemas baseados em conhecimento[31] (SBCs). Segundo Rezende (2003), os SBCs são apontados como um requisito indispensável na resolução de problemas complexos e se diferenciam dos tradicionais sistemas de processamento de dados na forma em que os dados são organizados, armazenados e manipulados, conforme pode ser conferido na Tabela 4.

Sistemas Convencionais	Sistemas Baseados em Conhecimento
Estrutura de dados	Representação do Conhecimento
Dados e Relações entre Dados	Conceitos, Relações entre Conceitos e Regras
Tipicamente usa algoritmos determinísticos	Busca Heurística
Conhecimento embutido no código do programa	Conhecimento representado explicitamente e separado do programa que o manipula e interpreta
Explicação do raciocínio é difícil	Podem e devem explicar seu raciocínio

Tabela 4 – Diferenças entre sistemas convencionais e sistemas baseados em conhecimento
Fonte: Rezende (2003, p.18)

A diferença em destaque aqui é relativa à forma em que os dados são organizados. Nos sistemas convencionais, os dados são organizados através de estruturas de dados, que

[30] "In AI, knowledge in computer systems is thought of as something that is explicitly represented and operated on by inference processes."

[31] Pesquisados desde a década de 70 pela comunidade de IA, juntamente com os sistemas especialistas. Estes, por sua vez, foram e ainda são dedicados em áreas que demandam tomadas de decisões de especialistas como a área de diagnóstico médico (RUSSELL e NORVIG, 2004, p.24).

representam os dados e suas relações. Os modelos para bancos de dados relacionais e os modelos orientados a objetos se enquadram nesses tipos de sistemas, conforme serão vistos respectivamente nas seções 2.3.2.1 e 2.3.2.2. Já nos sistemas baseados em conhecimento, os dados são organizados por meio de representação de conhecimento, que explicitam os conceitos do domínio do problema, suas relações e as regras de inferência, originando uma ontologia. Um SBC utiliza uma base de conhecimento, na qual se pode construir sentenças em uma linguagem de representação, modelando o problema que se deseja resolver. A ontologia, nesse caso, está associada à base de conhecimento no sentido de explicitar os conceitos e as relações entre conceitos, buscando representar todo o conhecimento relevante para a resolução do problema, além de apresentar definições sobre os significados dos conceitos e relações. Além da base de conhecimento, um SBC[32] possui um módulo separado da base, conhecido como mecanismo de inferência, capaz de realizar inferências sobre a base de conhecimento e obter conclusões a partir desse conhecimento. Baader et al. (1992, p.2) enfatiza que os serviços de inferência devem permitir responder algumas questões sobre a base que podem ajudar no raciocínio e resolução de problemas de um dado domínio como: a) o significado de um conceito é significativo para todos? (satisfatoriedade); b) onde exatamente o conceito está situado na hierarquia de conceitos? (classificação); c) quais fatos são dedutíveis a partir do conhecimento? (instanciação); d) de quais conceitos um objeto é instância? (compreensão); e) quais são as instâncias de um determinado conceito? (recuperação).

Contudo, para que a base de conhecimento possa ser criada e o mecanismo de inferência possa processar eficientemente a base torna-se necessária a adoção de métodos e técnicas para a construção de um SBC, visto não ser uma atividade trivial (REZENDE, 2003). Para tal, a área de Engenharia do Conhecimento emprega processos[33] constituídos de fases que vão desde a identificação da classe do problema que o SBC deverá resolver até a implementação da base em uma linguagem de representação. As técnicas de representação de conhecimento freqüentemente usadas são: as redes semânticas apresentadas na seção 2.3.1.1; os *frames* apresentados na seção 2.3.1.2; e a lógica apresentada na seção 2.3.1.3. As ontologias, apresentadas na seção 2.3.3, são também uma técnica de representação de conhecimento que consiste em organizar categorias e objetos em uma taxonomia através de

---

[32] Informações adicionais sobre a estrutura de um SBC podem ser encontradas em Rezende (2003, p.23). No que tange às ontologias apenas os componentes base de conhecimento e mecanismo de inferências são relevantes de serem apresentados no contexto dessa pesquisa.

[33] A apresentação de um possível processo é feita na seção sobre processo de construção de ontologias, tendo em vista que tal processo utiliza métodos e técnicas da Engenharia de Conhecimento.

declarações lógicas como a lógica de primeira ordem (RUSSELL e NORVIG, 2004, p.312). A importância do estudo de ontologias para a IA pode ser constatado pelo surgimento de uma nova área de trabalho conhecida como Engenharia ontológica, um ramo da Engenharia de Conhecimento, que se preocupa com o processo de construção de ontologias genéricas para fins de reuso (GRUBER, 1993) em diferentes bases de conhecimento.

2.3.1.1 Redes Semânticas

A utilização de nodos e arcos para a representação de conhecimento foi proposta na década de 60 por Quilliam[34], que propôs um modelo computacional da memória humana denominado *memória semântica*. A partir daí as redes semânticas foram sendo desenvolvidas na área de IA como uma forma de capturar o conhecimento de um domínio.

Rezende (2003) define uma rede semântica como um *grafo rotulado*[35] formado por: a) objetos que podem ser indivíduos, coisas e conceitos de um domínio, também denominados nós e b) um esquema de arcos ou arestas, também denominado *links*. A Figura 11 mostra um exemplo de uma rede semântica composta de categorias (Mamíferos, Pessoas, Pessoas Femininas e Pessoas Masculinas) e nomes de objetos (Maria e João) representados por elipses e as relações entre eles representadas por arcos rotulados. Russel e Norvig (2004) afirmam que a organização de objetos em categorias é uma atividade importante da representação do conhecimento, pois serve para organizar e simplificar a base de conhecimento por herança. Assim, como na orientação a objetos, as redes semânticas podem produzir dois tipos de relações: "é um" (*is a*), na qual as relações encontram-se numa taxonomia hierárquica; e "parte-de" (*part-of*), na qual um objeto é componente de outro.

[34] Informação obtida em http://www.das.ufsc.br/gia/softcomp/node27.html
[35] Um grafo é dito ser rotulado em arestas quando a cada aresta estiver associado um rótulo (uma etiqueta).

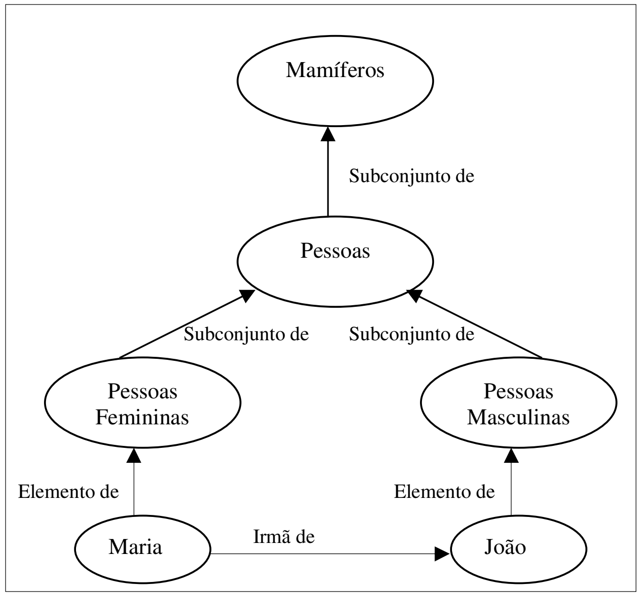


Figura 11 – Uma rede semântica com dois objetos e quatro categorias.
Fonte: adaptado de Russell e Norvig (2004, p.340)

Uma das vantagens desse instrumento de representação de conhecimento, apontada por Rezende (2003), é o poder que os mecanismos de inferência têm de derivar as propriedades de objetos mais gerais para os mais específicos (procedimento esse chamado de herança de propriedades). No exemplo da Figura 11, o objeto Maria recebe as propriedades da categoria Pessoas Femininas da qual faz parte, que por sua vez herda as propriedades das categorias Pessoas e Mamíferos. Russel e Norvig (2004) complementam dizendo que as redes semânticas oferecem gráficos que ajudam na visualização da base de conhecimento e algoritmos eficientes para dedução de propriedades de objetos de acordo com suas categorias. Entretanto, as representações gráficas impõem limitações expressivas pela falta de semântica formal que podem restringir o uso da linguagem (DONINI, LENZERINI e NARDI, 1997; REZENDE, 2003).

Os fundamentos das redes semânticas foram direcionados para alguns modelos que buscam organizar e representar conhecimento como os mapas conceituais (NOVAK e CAÑAS, 2008), que utilizam *links* para definir relacionamentos entre conceitos (também chamados de unidades semânticas) e a linguagem de modelagem unificada do paradigma orientado a objetos (BOOCH, JACOBSON e RUMBAUGH, 2006), que representa os diagramas de classes e de objetos através de associações semânticas entre eles. Algumas linguagens para descrição de recursos na Web (compõem a arquitetura da Web Semântica) utilizam um modelo de grafo, cuja fundamentação provém das redes semânticas. O modelo descreve os recursos e os objetos como nós e os predicados como arcos que ligam os nós. A modelagem unificada e alguns modelos de representação de conhecimento na Web serão

apresentados mais adiante nesse trabalho, tendo em vista a relação dos mesmos com as ontologias.

2.3.1.2 *Frames*

O conceito de *Frames* surgiu na década de 70 a partir do aparecimento da teoria dos *Frames*, uma derivação das redes semânticas, como uma forma de representar o conhecimento de um objeto através da observação visual do mesmo. Segundo Maida (1987)[36] e Minsky (1975)[37] apud Resende (2003), *Frame* é um termo usado para designar um agrupamento de conhecimento relevante a uma coisa, um indivíduo, uma situação ou um conceito. Possui um nome para identificar o conceito por ele definido e um conjunto de atributos denominados *slots*, que contêm valores ou ponteiros para valores. Um *slot* também possui um nome e é constituído de atributos, denominados *facetas*. As informações encontradas nas facetas explicitam os valores que o *slot* pode assumir ou a maneira de calcular ou deduzir o seu valor através de procedimentos (similar ao método da orientação a objetos). A Tabela 5 apresenta uma representação de um *Frame* Mobília.

Slots	Valores associados (facetas)
Especialização	Mobília
Número de cantos	Número inteiro (iniciando com 4)
Estilo de base	madeira
Número de cadeiras	2, 4 ou 6

Tabela 5 – Representação do *Frame* Mobília
Fonte: o próprio autor

Uma característica importante nesse tipo de instrumento de representação é a herança de propriedades, na qual uma classe especializada (especialização do frame-pai) pode herdar todas as propriedades da classe geral (generalização do frame-filho), permitindo a distribuição da informação sem duplicação (similar à estrutura hierárquica da orientação a objetos).

Assim como no modelo orientado a objetos, os *frames* possibilitam a simplificação de código e um sistema de fácil leitura e manutenção, porém não possuem semântica formal capaz de realizar deduções eficientes (BAADER, HORROCKS e SATTLER, 2003). Atualmente, os sistemas utilizam mais a orientação a objetos, entretanto existem muitos

---
[36] MAIDA, A.S. *Frame Theory*. John Wiley & Sons. 1987
[37] MINSKY, M.A. *Framework for Representing Knowledge*. New York: McGraw-Hill. 1975.

sistemas legados[38] utilizando as técnicas de *frames* (REZENDE, 2003). Algumas linguagens de representação de ontologias também incorporam a teoria dos *frames* em seus construtores como a Flogic - *Frame Logic* (KIFER, LAUSEN e WU, 1995), que integra *frames* com cálculo de predicados de primeira ordem; a RDF – *Resource Descrption Framework* (LASSILA e SWICK, 1999), que captura o formato e propriedades dos objetos na Web, além de especificar a semântica dos dados através de objetos e relacionamentos; e a OIL – *Ontology Interchange Language* (FENSEL et al., 2001), que contém as principais construções existentes em *frames* e em lógica descritiva. Além de linguagens, algumas ferramentas para construção de ontologias como o *Protégé – Frames*[39] utilizam *Frames* para especificar objetos que irão compor às ontologias.

2.3.1.3 Declarações Lógicas

O estudo sobre *lógica* e *regras de inferências* é de extrema importância para a área de Engenharia ontológica, visto que o conteúdo teórico das ontologias necessita de uma linguagem de representação (CHANDRASEKARAN, JOHNSON e BENJAMINS, 1999, p.21) para que posteriormente possa ser manipulado por um mecanismo de inferência. A intenção desta seção é fornecer uma visão teórica sobre alguns formalismos empregados, por exemplo, em metodologias para construção de ontologias como é o caso da lógica de primeira ordem utilizada no projeto *Toronto Virtual Enterprise* – Tove (GRUNINGER e FOX, 1995) e da lógica descritiva, que fundamenta a base formal de algumas linguagens para construção de ontologias no contexto da Web (BAADER, HORROCKS e SATTLER, 2003; BREITMAN, 2005).

Existem vários tipos de lógicas usadas para a realização de *dedução automática*[40]: a lógica proposicional, a lógica de primeira ordem e a lógica descritiva. A **lógica proposicional** é a mais simples por que se baseia apenas na existência de constantes (P, Q, R, etc) e no uso de conectivos lógicos: negação (¬), conjunção (^), disjunção (∨), implicação (→) e bicondicional – se e somente se (↔).

---

[38] Termo utilizado em referência aos sistemas de informações de organizações que, apesar de serem bastante antigos, fornecem serviços essenciais.
[39] Informações em http://protege.stanford.edu/overview/protege-frames.html.
[40] Refere-se ao comportamento de qualquer programa de computador que realizada inferências dedutivas a partir das leis da lógica matemática. Esta, por sua vez, possui várias regras sintáticas de dedução a partir das expressões da linguagem (REZENDE, 2003, p.30).

A semântica da lógica proposicional deve especificar como calcular o valor verdade de qualquer sentença, dado um modelo. Todas as sentenças são construídas a partir das sentenças atômicas (constantes P, Q) e dos conectivos lógicos. As regras para cada conectivo lógico podem ser resumidas em uma tabela-verdade[41], ilustrada na Tabela 6, que especifica o valor verdade de uma sentença complexa para cada atribuição possível de valores-verdade a seus componentes (RUSSELL e NORVIG, 2004).

P	Q	P ∧ Q	P ∨ Q	P → Q	P ↔ Q	¬P
V	V	V	V	V	V	F
V	F	F	V	F	F	F
F	V	F	V	V	F	V
F	F	F	F	V	V	V

Tabela 6 – Tabelas-verdade referente aos cinco conectivos lógicos
Fonte: adaptado de Russell e Norvig (2004, p.201)

No entanto, a lógica proposicional apresenta várias limitações, principalmente no que tange à expressividade (REZENDE, 2003). Observa-se, conforme o exemplo abaixo, que existem tipos de argumentos que, apesar de válidos, não é possível justificá-los com os recursos do cálculo proposicional.

Todos os humanos são racionais.
Alguns animais são humanos.
Logo, alguns animais são racionais.

As expressões "*todo*" e "*alguns*", por exemplo, não são contempladas na linguagem do cálculo proposicional. Isso seria resolvido através da introdução de novos símbolos, obtendo-se a linguagem do cálculo de predicados de primeira ordem apresentada a seguir.

A **lógica de primeira ordem** representa o mundo em termos de objetos e predicados desses objetos (propriedades dos *objetos* e *relação* entre eles), além de conectivos lógicos e quantificadores, sendo assim capaz de capturar uma boa idéia sobre o mundo (RUSSELL e NORVIG, 2004). Em relação aos quantificadores, a lógica de primeira ordem possui dois: i) quantificador *universal* (∀) indicando para *todo* ou para *todos* os objeto(s) e ii) quantificador *existencial* (∃) indicando declaração sobre *algum* objeto.

---

[41] Para verificar se uma sentença é verdadeira ou falsa: o valor de *P v Q* quando P é verdadeira e Q é falsa. Tome como ponto de partida as colunas P e Q a esquerda da tabela; procure a linha em que P é verdadeira e Q é falsa (segunda linha). Em seguida, observe nessa linha sob a coluna *P v Q* o resultado *verdadeiro*.

Uma sentença em lógica de primeira ordem anuncia um fato e é formada a partir de um símbolo de predicado, seguido por uma lista de termos[42] entre parênteses. A Figura 12 apresenta duas sentenças representando um possível cenário em uma agência de viagem. Na primeira sentença, *despesa_máxima* é o símbolo de predicado e *(c, b)* são símbolos de constantes ou termos, significando que c e b têm a propriedade *despesa_máxima*. Os símbolos ∃xy significam que alguma variável x como y tem uma propriedade, ou ainda, que existe no mínimo um objeto do universo considerado que tem uma propriedade. Na segunda sentença, os símbolos ∀$x significam que as propriedades envolvendo a variável x valem para todo x, ou ainda, que todos os objetos do universo considerado têm a propriedade em questão.

∃$x $y (destino($x) ∧ informação_de_viagem($y) ∧ quer_viajar(c,$y) ∧ idade(c,a) ∧ preferencias (c,e) ∧ despesa_maxima(c,b) ∧ tem_destino($y,$x)) ∀$x (viajante($x) ∧ (∃$y integer($y) ∧ idade($x,$y) ∧ ($y < 30)) ↔ viajante_jovem($x))

Figura 12 – Declarações em lógica de primeira ordem
Fonte: adaptado de Fernández-Lopez, Gomez-Perez e Corcho (2004, p.123)

Dessa forma, os axiomas definidos na Figura 12 podem responder a questões através do cálculo de predicados, como por exemplo, na primeira sentença: considerando as preferências de um viajante (viagem por cultura, viagem para as montanhas, viagem para a praia, etc) e algumas restrições (econômica ou sobre a própria viagem), *quais destinos são mais apropriados*? Onde as constantes significam: *c* é o viajante, *b* é a quantidade máxima que o viajante pode gastar, *a* é a idade do viajante, *e* são suas preferências; e as variáveis x e y significam respectivamente: destino da viagem e informação sobre a viagem. Uma outra questão poderia ser respondida: considerando um viajante jovem com um orçamento para alojamentos, *quais tipos de alojamentos estão disponíveis*? O axioma da segunda sentença expressa que o viajante (variável x) é considerado jovem se e somente se a sua idade for menor ou igual a 29 anos.

O cálculo de predicados seria então uma forma de raciocinar sobre predicados, quantificadores, conectivos lógicos e regras de inferência acerca de objetos presentes no universo considerado ou domínio.

[42] Um termo é uma expressão lógica que se refere a um objeto que pode ser símbolos de constantes ou variáveis. As variáveis representam objetos que não estão identificados no universo considerado, exemplo xy; já as constantes representam objetos identificados no universo, exemplo c, b.

Segundo Baader et al. (1992, p.1), "um mecanismo básico de processamento e representação de conhecimento humano é a divisão do mundo em classes ou conceitos [...] que geralmente são especificados por uma estrutura hierárquica" [43]. Russel e Norvig (2004, p.312) complementam assinalando que apesar da interação com o mundo ocorrer no nível de objetos individuais, uma grande parte do raciocínio tem lugar no nível de categorias, possibilitando fazer prognósticos sobre objetos, uma vez que eles estão classificados em suas devidas classes. A categoria pode deduzir a presença de certos objetos a partir da entrada de percepções como, por exemplo, deduzir que um objeto é uma maçã e que seria útil a uma salada de fruta a partir do conhecimento de seu formato, da casca vermelha, etc. Nesse sentido, surge na década de 80 a família de **lógica descritiva**, também conhecida como **lógica terminológica** (BAADER et al., 1992; DONINI, LENZERINI e NARDI, 1997), como uma possibilidade de representar conhecimento através da descrição formal de sua estrutura taxonômica (também chamada de terminologia) e fornecer um mecanismo de inferência especializado sobre tal estrutura (DONINI, LENZERINI e NARDI, 1997). A família de lógica descritiva é um subconjunto da lógica de predicados, tendo em vista que a notação da lógica de primeira ordem foi especificada para facilitar a tarefa de descrever objetos, conforme foi visto. Já as lógicas descritivas surgiram na perspectiva de tornar mais fácil a descrição de definições e propriedades de categorias ou classes (RUSSELL e NORVIG, 2004).

A característica da lógica descritiva (LD) em possuir uma semântica formal baseada em lógica veio a suprir as limitações (representação de conhecimento livre de contradições e ambigüidades) de seus predecessores como as redes semânticas e os *frames* (DONINI, LENZERINI e NARDI, 1997; BAADER, HORROCKS e SATTLER, 2003; CALVANESE e GIACOMO, 2003).

Nas linguagens terminológicas, *classes* são representadas por *conceitos* (os quais são usados para especificar um grupo de indivíduos) e *papéis* (os quais são relações binárias usadas para especificar propriedades ou atributos). Tais linguagens são baseadas em *primitivas* ou *axiomas terminológicos*, os quais são símbolos de um *alfabeto*. As primitivas são classificadas em conceitos atômicos ou nomes de conceitos, papéis atômicos ou nomes de papéis, nomes de atributos, nomes de indivíduos e nomes de objetos (BAADER et al., 1992, p.3).

---

[43] "One of the basic mechanisms of human knowledge representation and processing is the division of the world into classes or concepts […] which usually are given with a hierarchical structure."

Para a descrição formal da taxonomia na base de conhecimento, Donini, Lenzerini e Nardi (1997) elucidam a distinção entre *conhecimento extensional* e *conhecimento intensional* do domínio, nomeados[44] ABox (*assertional part*) e TBox (*terminological part*), respectivamente. Segundo os autores, o *conhecimento extensional* consiste no conhecimento específico sobre indivíduos do domínio. Já o *conhecimento intensional* está relacionado ao conhecimento geral sobre o domínio do problema, representado por uma taxonomia, a qual é construída através de declarações sobre propriedades dos conceitos e relações.

A Figura 13 ilustra a arquitetura de um sistema baseado em lógica descritiva. A base de conhecimento é dividida em conhecimento da terminologia e conhecimento sobre os objetos. A aplicação (interface responsável pela interação com o SBC) solicita informações ao SBC e, por meio do mecanismo de inferência, a linguagem de representação (nesse caso a lógica descritiva) usada na base de conhecimento é processada e as conclusões direcionadas à aplicação.

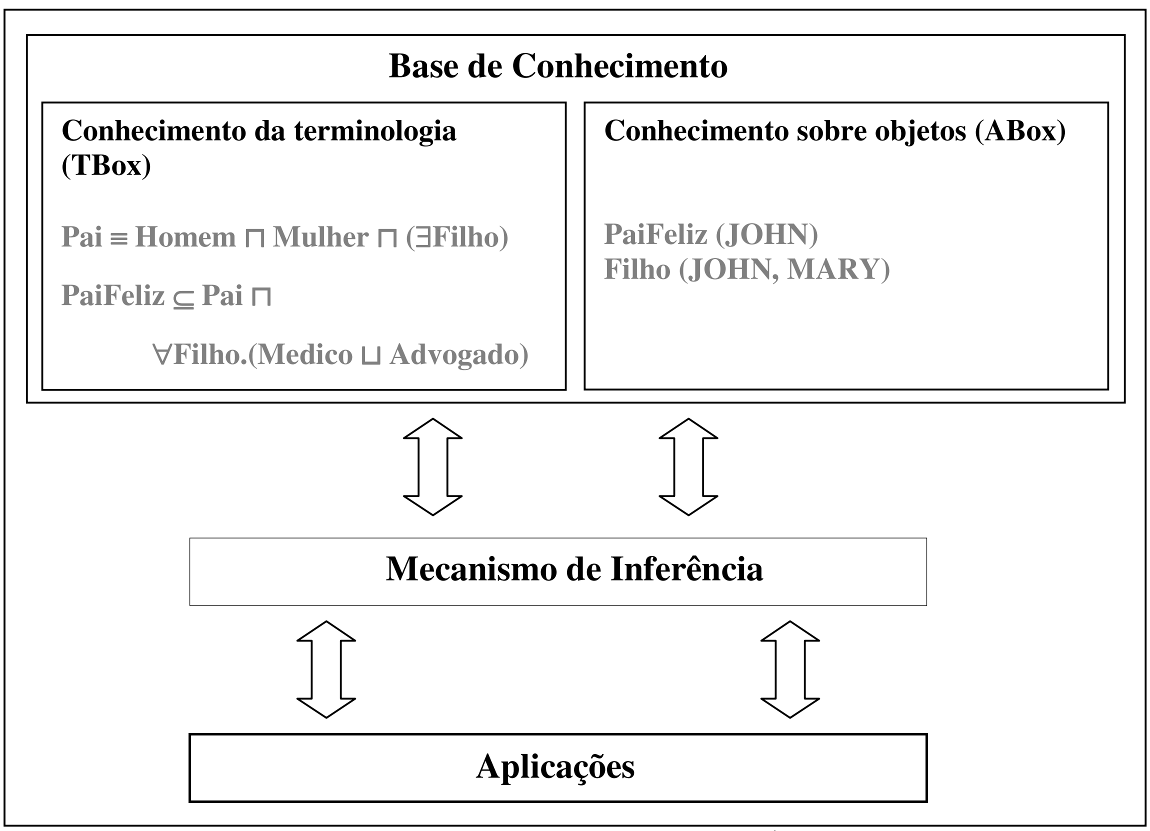


Figura 13 – Arquitetura de um sistema baseado em lógica descritiva
Fonte: adaptado de Calvanese e Giacomo (2003, p.6)

[44] Essa nomenclatura é originada do sistema KL-ONE (BRANCHMAN e SCHMOLZE, 1985), no qual foi pioneiro na caracterização lógica para a interpretação de objetos, de classes e de relações.

Além de conceitos e papéis atômicos, as lógicas descritivas permitem aos seus usuários construir expressões mais complexas de conceitos e papéis através dos axiomas terminológicos. Essas expressões são obtidas pela aplicação de construtores[45] sobre os axiomas terminológicos, definindo novos conceitos e papéis. Os exemplos a seguir, adaptados de Baader, Horrocks e Sattler (2003), ilustram alguns construtores típicos da lógica descritiva usados para definir conceitos.

- "*Um homem é casado com uma médica e possui no mínimo cinco filhos, sendo que todos eles são professores*". Tal conceito pode ser expresso em LD como se segue: Homem ⊓ ¬Mulher ⊓ ∃casado.Medico ⊓ (≥5 temFilho) ⊓ ∀temFilho.Professor. Nessa descrição, um individuo declarado Bob, por exemplo, pertence a *casado.Medico* se e somente se existir um individuo que seja *casado* com Bob (está relacionado a Bob por meio do papel *casado*) e seja uma médica (pertence ao conceito *Medico*). Da mesma forma, Bob pertence a *(>=5 temFilho*) se e somente se ele possuir no mínimo 5 filhos, além de pertencer a ∀*temFilho.Professor* se e somente se todos os seus filhos (todos os indivíduos relacionados a Bob por meio do papel *temFilho*) forem professores.
- "A*penas seres humanos podem ter crianças humanas*". A terminologia poderia ser declarada assim: ∃temFilho.Humano ⊆ Humano
- Ainda, uma declaração formal pode ser usada para indicar propriedades de indivíduos: HomemFeliz (Bob), temFilho (Bob, Mary) indica que Bob pertence ao conceito HomemFeliz e que Mary é uma de suas filhas.

A lógica descritiva fornece várias capacidades de inferências que deduzem o conhecimento presente na base (consiste de um conjunto de declarações e um conjunto de axiomas terminológicos). O algoritmo denominado subsunção (do inglês, *subsumption*) verifica se um conceito é um subconjunto de outro pela comparação de suas definições, e ainda verifica se um indivíduo (objeto) pertence a um conceito. Alguns exemplos são descritos a seguir.

---

[45] Construtores podem ser lógicos como de conjunção ou interseção (⊓), negação ou complemento (¬) e disjunção ou união (⊔). E ainda construtores de restrição existencial (∃), restrição de valor (∀), restrição de quantidade (≥) e restrição de hierarquia de papéis (⊆). O construtor C ≡ D representa uma abreviação para C ⊆ D, D ⊆ C (classes equivalentes).

- C é classificado em D se e somente se todas as instâncias de C são necessariamente instâncias de D. Por exemplo, o conceito de HomemFeliz descrito anteriormente: HomemFeliz $\subseteq$ ∃temFilho.Professor. HomemFeliz é classificado em ∃temFilho.Professor desde que as instâncias de HomemFeliz tenham no mínimo cinco filhos, e todos os filhos sejam professores.
- Um indivíduo é considerado instância da descrição do conceito C se e somente se este indivíduo for sempre interpretado como um elemento de C. Nas declarações feitas anteriormente, **Mary é uma instância de Professor**, o algoritmo de consistência determina se a base de conhecimento não é contraditória como no exemplo: se for adicionado na base ¬Professor (Mary), o algoritmo faz uma verificação nas declarações juntamente com a definição do conceito HomemFeliz e constata uma inconsistência.

Durante o final da década de 90, pesquisadores exploraram a idéia de como a representação de conhecimento da inteligência artificial poderia ser utilizada na Web. Baader, Horrocks e Sattler (2003) inserem a linguagem DARPA *Agent Markup Language* + *Ontology Inference Layer* - DAML+OIL, que tem sua fundamentação na lógica descritiva, no contexto da Web Semântica. Os autores ponderam sobre a capacidade da linguagem de definir, integrar e suportar ontologias de forma a proporcionar um entendimento comum sobre os conceitos usados nas páginas Web. Posteriormente, surgiu uma linguagem derivada da DAML+OIL, a *Ontology Web Language* – OWL, com forte impacto na Web Semântica (CARDOSO, 2007).

As linguagens de representação de ontologias, destacando a OWL, serão exploradas na seção 2.3.3.4, que destaca as ontologias como instrumento de representação de conhecimento.

2.3.2 Modelagem de dados

Dentre as principais contribuições da Ciência da Computação no que tange à representação de conhecimento destacam-se os modelos de representação associados à modelagem de dados, mais especificamente o Modelo Orientado a Objetos (RUMBAUGH e BLAHA, 2006; BOOCH, JACOBSON e RUMBAUGH, 2006), e o Modelo Entidade-Relacionamento (CODD, 1970; CHEN, 1976). A Ontologia Formal (GUARINO, 1998; GRUBER, 1993a) também está associada à questão de modelagem, porém o modelo repensa as possibilidades representacionais e de organização de domínios de conhecimento (CAMPOS, 2004) que serão discutidas em seção específica.

Todas essas tentativas de representação utilizam de mecanismos de abstração para a construção de modelos. A abstração nesse caso constitui uma ferramenta poderosa no exercício de aquisição de conhecimento, no sentido de se compreender certa realidade. Para tal, é necessário selecionar os objetos ou entidades de maior relevância para o problema objeto de investigação. Constroem-se, assim, esquemas abstratos da realidade, nos quais as coisas são reduzidas a seus perfis mais convenientes (SAYÃO, 2001, p.82).

Nesta seção, dar-se-á destaque ao Modelo Relacional em conexão com o Modelo de Entidade e Relacionamento, e ao Modelo Orientado a Objetos, destacando a sua linguagem de modelagem unificada, conhecida como UML (*Unified Modeling Language*). Tais modelos buscam representar os principais processos envolvidos em um ambiente específico, servir de instrumento de comunicação e construir um sistema de informação automatizado. Foram selecionados nessa revisão de literatura por apresentarem similaridades com a estrutura de uma ontologia, e por tratarem de questões relacionadas à semântica envolvida na representação de seus conteúdos. Contudo, em relação à semântica, observa-se uma diferença entre os dois modelos e as ontologias. Segundo Campos (2004, p.25), os modelos de dados e de objetos encontram-se em um nível epistemológico, ou seja, estabelecem apenas significados particulares de estruturação. Já as ontologias encontram-se em um nível ontológico, possibilitando sistematizar o conhecimento com definições axiomáticas.

2.3.2.1 Modelo Relacional

O modelo Relacional[46], criado por Edgar Frank Codd em 1970, é um modelo de dados adequado a ser o modelo subjacente de um Sistema Gerenciador de Banco de Dados (SGBD), que se baseia no princípio em que todos os dados estão guardados em tabelas (ou, matematicamente falando, em relações).

O modelo dos Diagramas de Entidade e Relacionamento, também chamado de DER, foi introduzido por Peter Chen (CHEN, 1976) na década de 70 e refinado posteriormente por outros pesquisadores. Surgiu da necessidade dos projetistas de bancos de dados encontrarem

---

[46] Anteriormente a esse modelo de implementação, dois modelos já haviam sido propostos na década de 60: o modelo em rede e o modelo hierárquico. A organização do modelo em rede se dá por um conjunto arbitrário de gráficos, já o modelo hierárquico através de estrutura em árvores. Essa organização limita a flexibilidade do modelo de dados, podendo ocorrer problemas de inconsistência no momento da manipulação dos dados.

uma maneira eficaz de estruturar tabelas, através do processo de normalização[47], eliminando a redundância de dados e os problemas de atualização inconsistente nas tabelas.

Chen (1976) define que no modelo DER as entidades (objetos ou coisas) são classificadas em diferentes conjuntos de entidades (representados por retângulos) através da observação de propriedades comuns sobre as entidades pertencentes ao domínio. Cada entidade possui os seus atributos, representados por elipses, que são propriedades específicas para a descrição da entidade. A Figura 14 mostra, como exemplo, o conjunto de entidades *funcionário* com seus atributos.

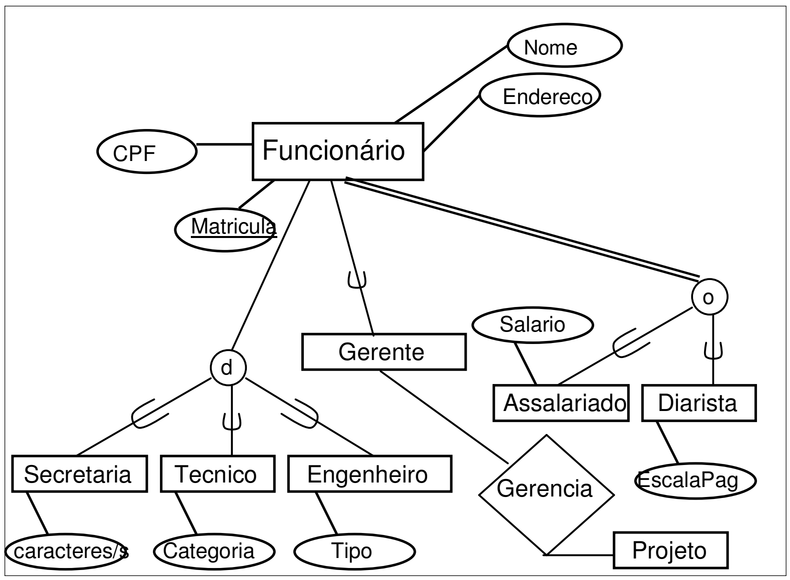


Figura 14 - Modelo de entidade e relacionamento estendido
Fonte: adaptado de Navathe e Elmasri (2000, p. 63)

O mecanismo de abstração advogado por Chen (1976, p.29) na organização e tratamento das entidades e dos relacionamentos é a abordagem *top-down*: "o modelo de entidade e relacionamento adota a abordagem *top-down*, utilizando a informação semântica para organizar dados em relações de entidade e relacionamento" [48]. Em tal abordagem, a análise semântica parte da formação de classes ou conjuntos de entidades e com base em alguma característica que distingue as entidades na classe, o modelo é estendido através de subclasses ou especializações, conforme mostra o exemplo da Figura 14: as entidades que são membros do tipo de entidade *funcionário* podem ser agrupadas ainda como *secretária*, *engenheiro*, *gerente*, *técnico*, *funcionário-assalariado* e *funcionário-diarista*. As razões

---

[47] Informações sobre normalização em bancos de dados relacionais podem ser conferidas em Silberschatz, Korth e Sudarshan (2006).
[48] "The entity-relationship model adopts a top-down approach, utilizing the semantic information to organize data in entity/relationship relations."

principais para incluir relacionamentos entre classes e especializações seriam no caso de certos atributos serem aplicados a algumas, mas não a todas as entidades da classe, como acontece na classe *funcionário*; e no caso de alguns tipos de relacionamentos somente poderem ter participação de entidades que são membros da subclasse, como acontece na relação entre a subclasse *gerente* e o conjunto de entidades *projeto*.

Uma entidade geralmente possui um atributo cujos valores são distintos para cada entidade no conjunto. Esse atributo é denominado chave primária[49] (ou do inglês *primary key*). Na Figura 14, o atributo *matricula* representa a restrição de unicidade do tipo de entidade *funcionário*. Tal restrição irá impor que as especializações *secretária*, *técnico* e *engenheiro*, por exemplo, herdem as propriedades (atributos e relacionamentos) da classe funcionário. O relacionamento por meio de herança faz parte da extensão do DER conhecida como *modelo de entidade e relacionamento estendido*, o qual apresenta pontos similares às ontologias, bem como ao modelo orientado a objetos no que diz respeito à estrutura formada por classes e especializações.

A propriedade de relacionamento é quem estabelece a associação entre uma ou várias entidades através de operações da álgebra relacional, fundamentada na teoria de conjuntos da matemática: união (propriedades disjuntas), interseção (propriedades sobrepostas), diferença (propriedades encontradas em um conjunto, mas não em outro) e produto cartesiano (composição de propriedades). A Figura 14 mostra uma situação disjunta (através do símbolo **d** - *disjunct***)** entre o conjunto de especializações *secretária*, *técnico* e *engenheiro* e sua classe *funcionário*. Tal restrição impõe ao modelo que uma entidade funcionário somente poderá fazer parte de uma das especializações. Já numa situação de sobreposto, uma entidade poderia fazer parte de duas ou mais especializações (utilizar-se-ia o símbolo **o** - *overlap*).

Algumas semelhanças entre o modelo entidade e relacionamento e as ontologias podem ser conferidas no tocante às restrições impostas aos dois modelos: estrutura de classes e especializações e associações entre entidades ou objetos. Ambos tratam das “coisas do mundo” e buscam um significado mais próximo do mundo real através da semântica embutida nas propriedades das “coisas”. No tocante às ontologias, o número de associações entre os objetos não é limitado como acontece no modelo de entidade e relacionamento. As linguagens de representação de ontologias são consideradas semanticamente expressivas para a criação

---

[49] Para fins de projetos de banco de dados torna-se fundamental a imposição de outra restrição para garantir a integridade do banco. É a referência da chave primária em um outro tipo de entidade (denominada chave estrangeira - do inglês *foreign key*) através da restrição de cardinalidade. A cardinalidade expressa o número de entidades às quais a outra entidade se relaciona através do conjunto de relacionamentos.

de relações - OIL (FENSEL et al., 2001), DAML+OIL (HORROCKS et al., 2001) e OWL (DEAN et al., 2003) - que dão a elas um poder semântico superior aos modelos para bancos de dados. No modelo relacional, os dados são disponibilizados de maneira estruturada, possibilitando uma busca precisa no banco de dados[50]. Já nas ontologias, a representação vai além de dados isolados, isto é, o instrumento lida com conhecimento em sua estrutura. Tal conhecimento pode ser representado por formalismos lógicos que fornecem aos fatos do mundo um aspecto semântico próximo da realidade em que se deseja representar. Esse diferencial semântico permite às ontologias utilizarem mecanismos de inferência para recuperar informações de maneira dedutiva e significativa para o usuário.

2.3.2.2 Modelo Orientado a Objetos

A modelagem orientada a objetos tem como meta identificar o melhor conjunto de objetos para descrever um sistema de *software* e reusar tais objetos quando necessário em outras aplicações (RUMBAUGH e BLAHA, 2006). Um objeto pode ser considerado um conceito, uma abstração ou coisa que representa, através de sua identidade, algum significado para uma aplicação.

Na modelagem orientada a objetos implementa-se um conjunto de classes que definem os objetos presentes no sistema de *software*. Uma classe representa um grupo de objetos semelhantes e que através da *especialização* e *generalização* ou, simplesmente, categorização, organiza os objetos por suas semelhanças e diferenças (RUMBAUGH e BLAHA, 2006). Desta maneira, surge o conceito de superclasse e subclasses. Cada classe determina o comportamento (definido nos métodos[51]) e estados possíveis (atributos) de seus objetos, assim como o relacionamento com outros objetos. Cada um desses objetos é chamado de instância de sua classe. Segundo Rumbaugh e Blaha (2006), a subclasse herda as características de sua superclasse. E acrescentam que cada instância de uma subclasse é também uma instância da superclasse. Por isso da denominação "é um", do inglês "*is a*", para o relacionamento de generalização. A Figura 15 ilustra uma situação de especialização e generalização entre a superclasse pessoa e suas subclasses cliente e empregado, que por sua vez especializa mais três subclasses: escriturário, caixa e secretária.

---

[50] Através de uma linguagem de banco de dados, a *Structure Query Language*, conhecida como SQL.

[51] São usados para mudar os valores dos atributos do objeto ou retornar valores de atributos de objetos selecionados. A mesma operação pode ser aplicada a diferentes tipos de objetos; em tal situação, um nome de operação pode ser referir a várias implementações distintas, mas logicamente representando a mesma tarefa. Esse procedimento é conhecido como polimorfismo (RUMBAUGH e BLAHA, 2006).

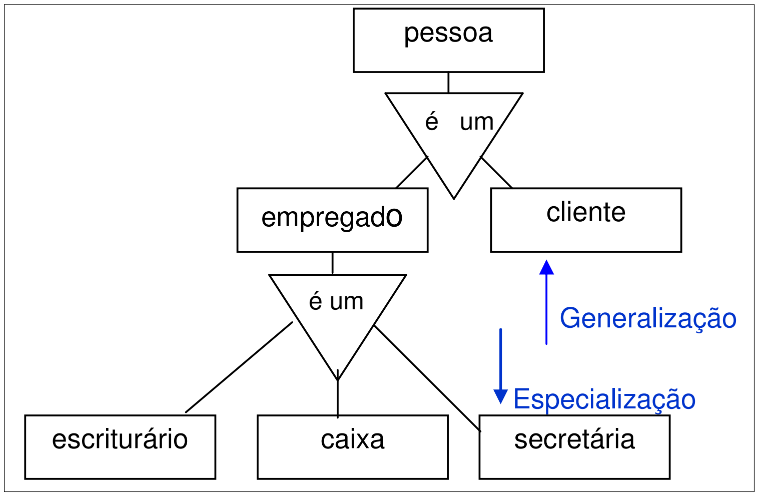


Figura 15 - Modelo representando uma herança entre a classe pessoa e suas subclasses
Fonte: adaptado de Silberschatz, Korth e Sudarshan (2006)

Outras formas de organizar ou classificar objetos podem ser feitas através da *agregação* e da *associação*. Na agregação há níveis de abstração do tipo "parte de" em relação à classe. Exemplo: *termostatos* e *compressores* são partes de *congeladores*. Na associação, uma classe tem atributos que associam características comuns a outras classes diferentes. Nessa classificação, certas idéias promovem lembranças de outras idéias relacionadas, como por exemplo: *montanhas* e *pistas* estão associadas a *esquiar*.

Salienta-se nesse ponto a similaridade entre as relações apresentadas e as relações entre conceitos observadas nos instrumentos discutidos nesse trabalho, ontologias e tesauros, por exemplo. As relações de especialização e generalização, de agregação e de associação são similares às utilizadas nos tesauros, conforme foi visto na seção 2.4.3.3.1. Nos tesauros, a especialização e a generalização são representadas respectivamente pelos símbolos TE (termo específico) e TG (termo genérico), a agregação é denominada relação hierárquica partitiva, e a associação é identificada pelo símbolo TR (termo relacionado). Nas ontologias, a especialização e a generalização são relações representadas através de hierarquias de classes, conhecidas como taxonomia. Tais relações, bem como as de agregação e de associação podem ser representadas através de uma linguagem lógica, como a lógica descritiva e a lógica de primeira ordem (USCHOLD, 1996). A lógica de primeira ordem e a lógica descritiva são vistas como possibilidades de tipos de deduções projetadas para auxiliar o usuário a definir um conjunto coerente de conceitos e objetos para descrever o domínio (CRANEFIELD e PURVIS, 1999). Além da similaridade entre as relações, as restrições e as deduções impostas sobre os valores dos atributos dos conceitos, na ontologia, são implementados no modelo orientado a objetos através das assinaturas das classes (representadas pelos métodos).

Na atividade de representação das classes, dos objetos e das relações entre esses objetos, a comunidade de orientação a objetos vem utilizando a linguagem de modelagem unificada, a UML, como técnica de modelagem de sistemas. Pesquisadores das áreas de ontologias e modelagem orientada a objetos (EVANS, 1998; EVANS et al., 1998; CRANEFIELD e PURVIS, 1999; OMG, 2003) vêm agregando esforços em trabalhos que utilizam a UML na especificação de ontologias. Isso se deve ao fato de que, na condução de projetos orientados a objetos, a UML é considerada a técnica de modelagem atualmente mais utilizada por sua facilidade de comunicação através de representação gráfica, e pelo fato de as ontologias possuírem uma semântica mais expressiva em relação à UML na representação de um domínio de conhecimento. Esse cenário é constatado por Campos (2004, p.31) que assinala que a ontologia "[...] utiliza-se da representação gráfica, como uma ferramenta para garantir um projeto lógico mais bem estruturado de um sistema e, para tanto, recorre geralmente à representação gráfica utilizada pelo modelo OO".

Na seção a seguir a UML é apresentada brevemente no sentido de elucidar possíveis contribuições para as ontologias na concepção de modelos formais e interativos à comunidade de usuários.

2.3.2.2.1 Linguagem de Modelagem Unificada

A UML é sucessora de um conjunto de métodos de análise e projetos baseados em objetos, criado pelos próprios autores da linguagem: Grady Booch, James Rumbaugh e Ivar Jacobson. A linguagem é adotada e recomendada pelo *Object Management Group* – OMG, um grupo responsável pelos padrões voltados ao paradigma de orientação a objetos. Por possuir uma notação gráfica simples e ser capaz de modelar aspectos estruturais e comportamentais de um sistema é amplamente consolidada na área de Engenharia de Software (PRESSMAN, 2002; SOMMERVILLE, 2007; BOOCH, JACOBSON e RUMBAUGH, 2006).

Furlan (1998, p.38) assinala que a UML é uma linguagem de modelagem e não uma metodologia, pois não explicita os procedimentos de uso da linguagem para construção de *softwares*, geralmente embutidos numa metodologia.

A UML possui vários diagramas (BOOCH, JACOBSON e RUMBAUGH, 2006). Os diagramas que modelam os aspectos estruturais são os diagramas de classes, diagramas de objetos, diagramas de componentes e os diagramas de desenvolvimento. Os diagramas de

colaboração, diagramas de seqüência, diagramas de casos de uso, diagramas de estados e os diagramas de atividades servem para modelar os aspectos comportamentais de um sistema. A Figura 16 mostra um exemplo de um diagrama de classes sobre um catálogo de CDs de música clássica.

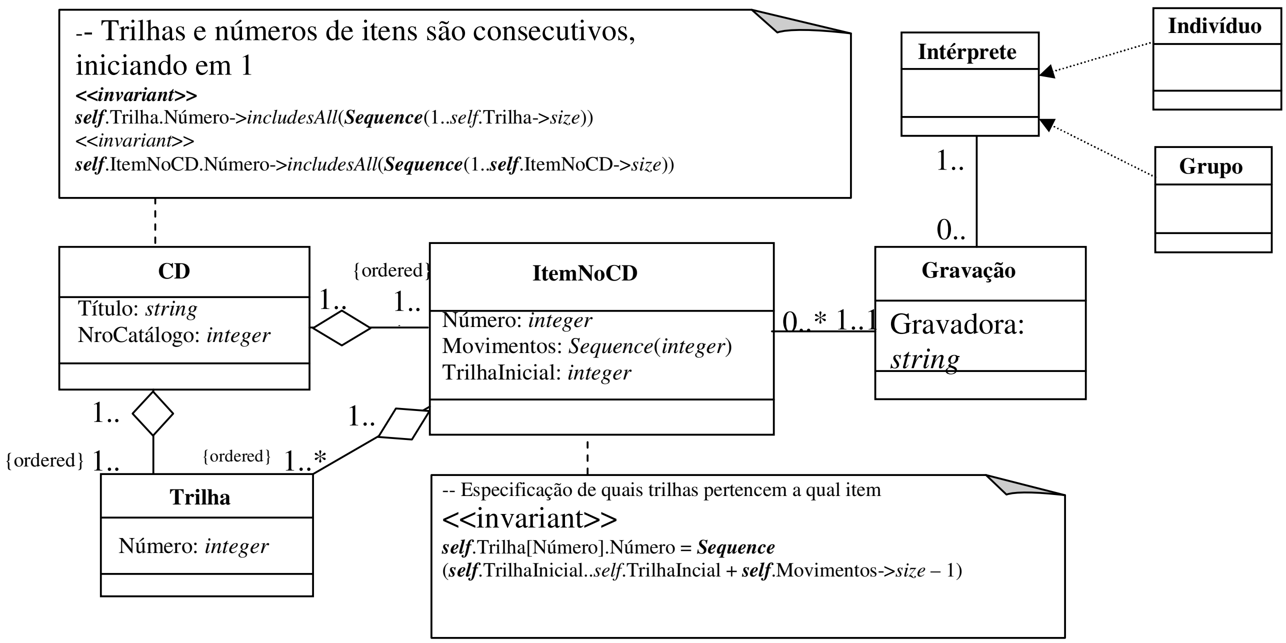


Figura 16 – Exemplo de Diagrama de Classe
Fonte: adaptado de Cranefield e Purvis (1999)

O diagrama de classes mostra um conjunto de classes, interfaces e relacionamentos que é responsável por modelar a parte estática do sistema. A semântica estática consiste em um conjunto de regras que definem diagramas bem formados dentro da linguagem. No exemplo da Figura 16 as classes *Indivíduo* e *Grupo* são especializações da classe *Intérprete*, relacionamento conhecido como generalização; outros tipos de relacionamentos são contemplados no diagrama de classes como o relacionamento de associação entre as classes *Gravação* e *ItemNoCD*, que representa uma relação estrutural entre duas classes indicando que estas se comunicam através e troca de mensagens; e o relacionamento de dependência entre a classe *ItemNoCD* e *CD*, chamado de relacionamento de agregação, cuja semântica é "parte de".

A percepção da relação do modelo orientado a objetos para com as ontologias torna-se clara na medida em que ambos buscam representar semanticamente um domínio de conhecimento. O primeiro modelo busca classificar o domínio através de classes e objetos, possibilitando o reuso dos mesmos em outras aplicações. Tais princípios também são

empregados nas ontologias (NOY e GUINNESS, 2001) quando concebidas para fins de modelagem de domínio.

Dada a similaridade entre os modelos orientados a objetos e as ontologias, a comunidade de orientação a objetos tem despertado para o uso de ontologias através da UML em conjunto com a sua linguagem de restrição de objetos denominada OCL - *Object Constraint Language* (ERIKSSON e PENKER, 1998). A OCL[52] foi desenvolvida para preencher a necessidade existente na UML de representar, formalmente, restrições adicionais sobre os objetos do modelo. A linguagem restringe valores de atributos e possíveis instâncias de um relacionamento, além de especificar procedimentos para verificação de integridades. A Figura 16 mostra dois exemplos de restrições em OCL vinculadas às classes *CD* e *ItemNoCd*. A especificação desta última restringe as possíveis instâncias no relacionamento de agregação entre *Trilhas* e *Itens*. A palavra-chave *self* indica uma instância; a palavra *sequence*, um tipo de lista ordenada e *size* é uma função que retorna o número de elementos de uma seqüência da lista ordenada.

A UML e a OCL ainda não possuem uma semântica formal. A OMG fornece uma descrição informal em linguagem natural (CRANEFIELD e PURVIS, 1999). Entretanto, pesquisadores vêm propondo diferentes formas semânticas para UML: modelo matemático direto (BREU et al, 1997); descrição utilizando a linguagem de especificação Z (EVANS et al, 1998); definição de uma semântica formal, bem como de regras de transformação dedutivas para provar que um diagrama é conseqüência de outro (EVANS, 1998); operações semânticas descrevendo como um modelo UML se envolve com outros elementos (OVERGAARD, 1998); e semântica para OCL nos diagramas de classes tem sido proposta por RICHTERS e GOGOLLA (1998).

Ressalta-se a importância da semântica na especificação de modelos para a representação de conhecimento, podendo ser alcançada através do uso de ontologias; e a facilidade na forma de representação gráfica proporcionada pela UML, viabilizando a interação entre comunidades de usuários e projetos de sistemas de informações. Iniciativas fundamentais na junção das comunidades de orientação a objetos e ontologias encontram-se num patamar evolutivo, como é o caso da especificação de ontologias DAML-OIL (LOCKHEED, 2000) e ontologias OWL (BROCKMANS e HAASE, 2006; OMG, 2003) em UML, no contexto da Web Semântica.

---

[52] Informações adicionais sobre as linguagens UML e OCL podem ser conferidas no sítio www.omg.org.

Na próxima seção, as ontologias são apresentadas com maior detalhe, de forma a elucidar as relações entre a representação de conhecimento e os modelos semânticos.

2.3.3 Ontologias e Representação do Conhecimento

A representação do conhecimento como campo de pesquisa na Inteligência Artificial vem concentrando seus esforços em técnicas de apreensão de conhecimento do domínio e explicitação do mesmo através dos sistemas baseados em conhecimento. As ontologias surgem como uma possibilidade na área de IA, especificamente na Engenharia ontológica, de especificar modelos formais, através de declarações lógicas, para representação de domínios de conhecimento, buscando tratar questões de natureza semântica envolvendo "as coisas" pertencentes ao domínio da aplicação, e viabilizar o tratamento da linguagem padronizada para que o conhecimento do domínio possa ser expresso e compartilhado.

Além da IA, outros campos do conhecimento têm-se interessado nos estudos sobre ontologias, como é o caso da Ciência da Informação. Neste campo, buscam-se alternativas para amenizar os problemas na comunicação entre sistemas de informações e seus usuários. Pode-se ter, através de ontologias, um ponto de partida na modelagem conceitual de determinado domínio, buscando um entendimento compartilhado do mesmo. Desta forma, a comunidade de usuários envolvida no domínio passa a ter uma linguagem comum através da qual o seu conhecimento possa ser expresso, facilitando a comunicação entre esta comunidade e o sistema. Falbo (1998) afirma que o emprego de ontologias pode facilitar a comunicação entre diversos usuários com visões diferentes do processo através de um vocabulário compartilhado do domínio de interesse.

Esta pesquisa busca contribuir nos estudos sobre ontologias construindo um arcabouço teórico sobre as mesmas. A seção 2.3.3.1 busca o esclarecimento do significado do termo "ontologia" em seu campo de origem, a Filosofia, bem como em outros campos que fazem uso do termo; a seção 2.3.3.2 apresenta as principais características das ontologias e tipos definidos; a seção 2.3.3.3 destaca as principais aplicações que fazem uso; a seção 2.3.3.4 enfatiza as principais linguagens de representação de ontologias, destacando o advento das linguagens de marcação e padrões de metadados para o contexto da web semântica, uma das aplicações mais recentes que faz uso de ontologias; e, finalmente, a seção 2.3.3.5 aborda processos de construção de ontologias mais comumente encontrados na literatura, fornecendo um panorama interdisciplinar de envolvimento de áreas como a Engenharia de Software e a Engenharia do Conhecimento no ciclo de vida das ontologias.

É válido ressaltar a importância dessa seção para a pesquisa, visto que o trabalho se propõe a fazer um estudo analítico de metodologias para construção de ontologias a fim de propor uma nova possibilidade metodológica para construção de tal instrumento. Nesse sentido, almeja-se que os aspectos teóricos elucidados nessa seção possam ajudar o leitor no entendimento sobre a temática em questão.

2.3.3.1 Origem e Definições

Ontologia como um ramo da filosofia remete ao "[...] estudo do Ser; a área da metafísica que se relaciona ao Ser ou essência das coisas, ou o Ser no sentido abstrato" (OXFORD ENGLISH DICTIONARY). Já na computação, uma ontologia é um artefato de *software* que tem utilizações específicas em ambientes computacionais (SMITH, 2004). A apropriação do termo "ontologia" da filosofia pela comunidade de computação (em especial pela comunidade de inteligência artificial) deve-se ao fato de as ontologias servirem como meio de *organização das coisas* passíveis de representação simbólica (representação formal). E, a partir da representação formal, possibilitar raciocínio dedutivo através de regras de inferências. Tal organização formal remete ao termo "ontologia formal" que Cocchiarella (1991)[53] apud Guarino (1995) define como "desenvolvimento sistemático, formal e axiomático da lógica de todas as formas e modos do Ser"[54]. Guarino (1995) complementa afirmando que o sentido formal é relacionado ao rigor imposto à *descrição das formas do Ser*, que em outras palavras significa dizer: expressividade semântica atribuída ao Ser.

No processo de *descrição das formas do Ser*, as distinções entre entidades (objetos físicos, eventos, regiões, etc) e entre meta categorias (conceitos, propriedades, qualidades, estados, papéis, partes, etc) tornam-se relevantes ao processo de modelagem do mundo (GUARINO, 1995) e vai ao encontro das categorias de Aristóteles (CORAZZON, 2008a) na ação de classificar as coisas para organizar o conhecimento do mundo.

Na Ciência da Computação os estudos sobre ontologias como artefato de *software* tiveram início na década 90, principalmente na Inteligência Artificial em pesquisas sobre representação do conhecimento (GRUBER, 1993; GRUBER, 1993a; GRUBER, 1994; GUARINO, 1995; GUARINO e GIARETA, 1995). O interesse sobre o assunto ontologias na Ciência da Informação (SOERGEL, 1997; SOERGEL, 1999; VICKERY, 1997) acontece

---

[53] COCCHIARELLA, N. B. 1991. Formal Ontology. In H.Burkhardt and B. Smith (eds.), *Handbook of Metaphysics and Ontology*. Philosophia Verlag, Munich: 640-647.

[54] "The systematic, formal, axiomatic development of the logic of all forms and modes of being."

também nesse período. Vickery (1997) foi um dos primeiros do campo da Biblioteconomia a dar atenção ao termo ontologia na Ciência da Informação.

Pesquisadores dos campos de Ciência da Computação e Ciência da Informação abordam o conceito de ontologia explicitando seu papel de estruturar um domínio de conhecimento e compartilhar tal conhecimento numa comunidade de interesse. Jurisica, Mylopoulos e Yu (1999) afirmam que as ontologias podem ser usadas como conhecimento comum de um domínio, viabilizando a comunicação entre uma comunidade de interesse. Para os autores, na perspectiva da Ciência da Computação e da Ciência da Informação, uma ontologia pode ser útil na organização e representação do conhecimento, tendo a tecnologia como apoio na viabilização de uma infra-estrutura para gerência de conhecimento.

Numa perspectiva interdisciplinar entre a Inteligência Artificial e a Filosofia, Chandrasekaran, Johnson e Benjamins (1999) definem que uma ontologia refere-se a conteúdo teórico sobre diversos objetos, a propriedades desses objetos e ao relacionamento entre objetos que são possíveis num domínio específico de conhecimento.

Gruber (1993, p.199) define ontologia como uma "especificação explícita de uma conceitualização". Tal definição é discutida por alguns autores no que diz respeito à noção do termo "conceitualização". Guarino e Giareta (1995) afirmam que o problema está em considerar a noção extensional ao definir "conceitualização". Segundo os autores, a noção extensional está relacionada a um estado particular das coisas e não representa a estrutura semântica de certa realidade ou **modelo pretendido**. Uma **conceitualização** seria assim, melhor definida a partir de uma noção intensional, na qual se considera a conceitualização como um conjunto de modelos semânticos os quais representam certa realidade. A Figura 17 apresenta uma possível abstração para uma conceitualização, a ontologia conceitual geral do mundo:

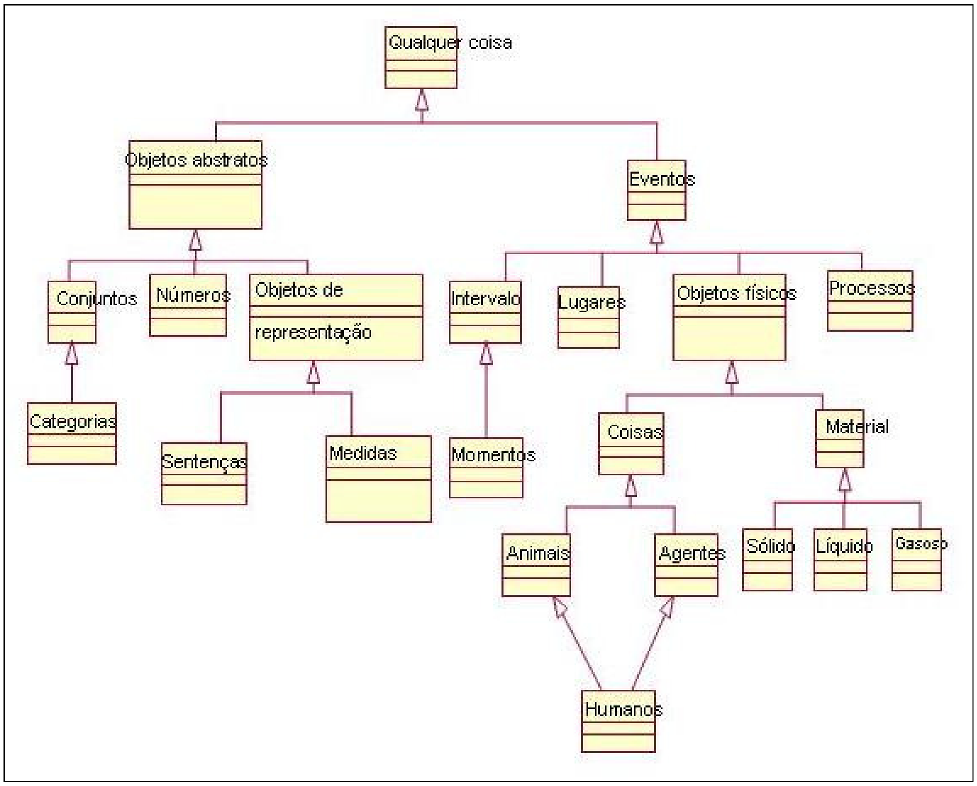


Figura 17 – Ontologia geral do mundo representada em UML
Fonte: adaptado de Russell e Norvig (2004, p.310)

A interpretação intensional é determinada pelo **compromisso ontológico** entre a **linguagem** do modelo pretendido (ou **vocabulário** pretendido) e a conceitualização. Nesse sentido, a conceitualização precisa se comprometer em usar um vocabulário comum ou compartilhado que seja coerente com a linguagem do modelo pretendido, tendo em vista que tal linguagem pode incluir axiomas que restringem a utilização do vocabulário. Desse modo, um conjunto de modelos constitui uma conceitualização, já os modelos são construídos através de uma linguagem de modelagem, especificada pela ontologia.

A conceitualização é assim, **independente** da linguagem do modelo, já a ontologia em si permanece **dependente** da linguagem, e de seus axiomas específicos.

Diante das explanações acima, é válido afirmar que uma ontologia possui uma conceitualização que a fundamenta e que a mesma ainda pode fundamentar diferentes modelos. As relações entre linguagem, conceitualização, compromisso ontológico e ontologia podem ser conferidas na Figura 18.

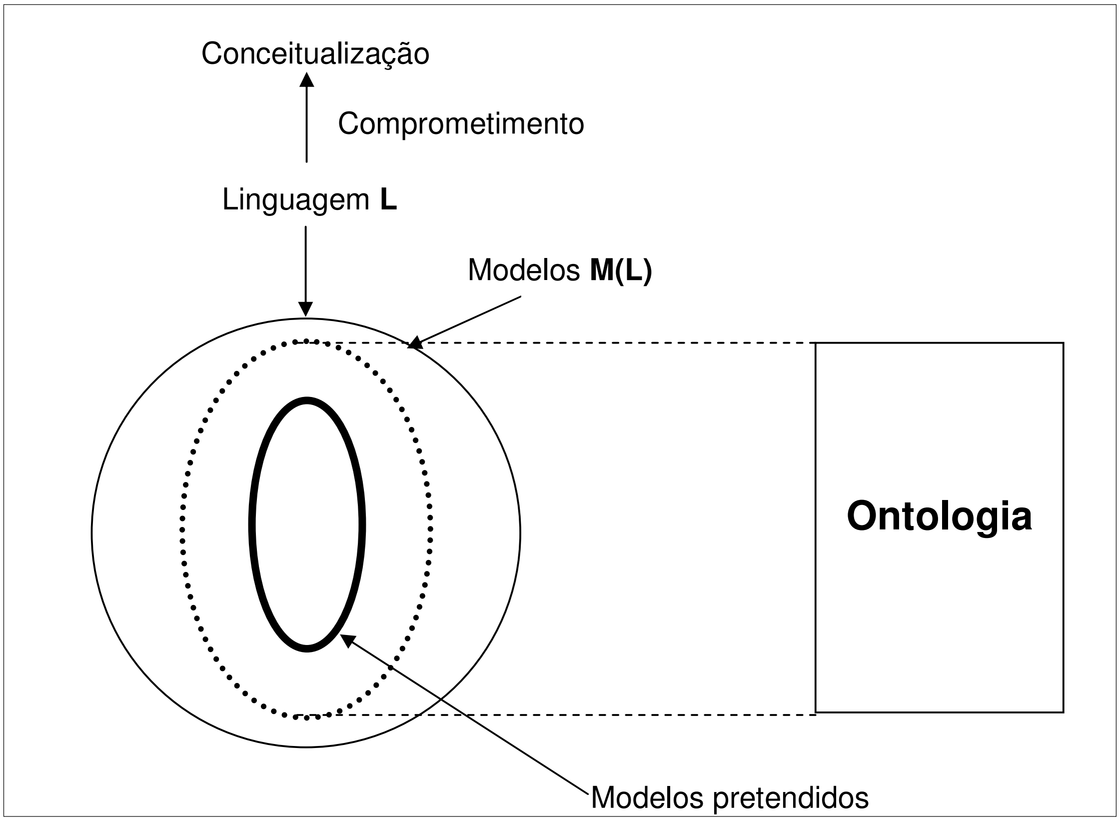


Figura 18 – Representação de uma ontologia e sua relação com a conceitualização
Fonte: adaptado de Guarino (1998, p.7)

Finalmente, para uma ontologia se aproximar de um modelo pretendido, conforme mostra a Figura 18, deve-se obter uma conceitualização com boa axiomatização, ou seja, obedecendo a regras formais que restringem a estrutura de parte da linguagem, conformando-a a realidade, mais um conjunto formal de conceitos e relações relevantes estabelecido de maneira consensual entre a comunidade envolvida. Dessa forma, Guarino (1998, p.5) aprimora a definição de Gruber descrevendo uma ontologia como sendo um artefato de engenharia, constituído por um vocabulário intensional utilizado para descrever certa realidade, em conjunto com pressupostos explícitos em forma de lógica de primeira ordem representando conceitos (também denominado predicado unário) e relações entre conceitos (também denominado predicado binário). Tal vocabulário intensional é descrito mediante uma aceitação, ou seja, um consenso no sentido das palavras que irão compô-lo.

Borst (1997, p.12) também apresenta uma definição muito aceita pela comunidade de ontologia: “uma especificação formal e explícita de uma **conceitualização compartilhada**”. Ou seja, um **conhecimento consensual** de um determinado domínio. Almeida e Bax (2003) explicam que “formal” significa legível para computadores; “especificação explícita” estaria

relacionada a conceitos, propriedades, axiomas explicitamente definidos; “compartilhado” seria conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real.

Para Uschold e Gruninger (1996), ontologias são termos usados para referenciar um conhecimento compartilhado em algum domínio de interesse, que pode ser usado como uma aplicação unificada para resolver problemas. Afirmam que uma ontologia necessariamente incluirá um vocabulário de termos e alguma especificação de seu significado. Esse vocabulário pode ser representado com distintos graus de formalismo: desde o altamente informal, utilizando-se linguagem natural, até o rigorosamente formal, representando-se os termos por meio de teoria lógica.

Pesquisadores da Ciência da Informação (SOERGEL, 1997; SOERGEL, 1999; VICKERY, 1997; GILCHRIST, 2003; FISCHER, 1998) apresentam similaridades entre instrumentos utilizados na biblioteconomia, como os tesauros, os dicionários e as taxonomias, e instrumentos utilizados na Inteligência Artificial, como as ontologias. As similaridades estão principalmente na forma de elaboração da estrutura desses instrumentos, que demanda a organização de conceitos em processos que incluem categorização e classificação de conceitos, definição das relações entre esses conceitos e tratamento da terminologia empregada nos conceitos e relações da estrutura.

O aporte metodológico e as bases teóricas consistentes encontradas na Ciência da Informação, bem como na Biblioteconomia, podem contribuir com métodos e técnicas na atividade de elaboração de classificações (RANGANATHAN, 1967), de definições (DAHLBERG, 1978) e de relações entre conceitos (WÜESTER, 1981[55] apud CAMPOS, 2001). Campos (2004, p.26) assinala que na atividade de definição está implícita uma classificação. Segundo a autora “Definir é evidenciar não somente quais são os elementos constitutivos do objeto e sua funcionalidade em dado contexto, mas também o que é o objeto. Desta forma está implícita uma estrutura classificatória”. E complementa dizendo que apesar de avançar em aspectos teóricos, a ontologia “[...] ainda está refém de uma forma de evidenciar os conceitos e suas relações [...]” (CAMPOS, 2004, p.31).

Soergel (1999) destaca a carência de comunicação entre as comunidades que lidam com estruturas conceituais, nas quais os esforços para construção de produtos são fragmentados, resultando em considerável reinvenção de conceitos, como por exemplo,

---

[55] WUESTER, E. L. L’étude scientifique qénérale de la terminologie, zone frontalière entre la linguistique, la logique, l’ontologie, L’informatique et les sciences des chose. In: RONDEAU, G. ; FELBER, E. (Org.). Textes choisis de Terminologie. Québec: GIRSTERM, 1981. p. 57-114.

classificação. Segundo o autor, uma ontologia "é uma classificação de categorias básicas" [56]. Sendo assim, as áreas que lidam com desenvolvimento de ontologias poderiam recorrer a técnicas utilizadas na Ciência da Informação para aproveitamento de metodologias já consolidadas no campo.

Vickery (1997) destaca questões presentes na pesquisa de ontologias que também são abordadas pela Biblioteconomia, como a categorização de conceitos – principio básico da teoria da classificação. E conclui que, apesar da similaridade evidente, os autores da Inteligência Artificial não referenciam trabalhos importantes da Biblioteconomia, como, por exemplo, os métodos e técnicas empregados por Lancaster (1986) na construção de vocabulários controlados direcionados a sistemas de recuperação de informação.

2.3.3.2 Características e Tipos de Ontologias

Os componentes básicos de uma ontologia podem ser encontrados em Gruber (1993a): a) **classes** – organização dos conceitos de um domínio, organizadas em uma taxonomia. As classes herdam as características de suas classes-pai, de forma similar ao modelo orientado a objetos; b) **relações** - representam o tipo de interação entre os conceitos de um domínio; c) **axiomas** - usados para restringir a interpretação e o uso dos termos envolvidos; d**) instâncias** - utilizadas para representar objetos específicos. No que diz respeito aos axiomas, a lógica descritiva pode vir a ser empregada na incorporação de instruções lógico-formais, que possibilitariam o uso de mecanismos de inferência (ALMEIDA e BAX, 2003). Isto possibilitaria, por exemplo, a classificação automática de um conceito na hierarquia, de forma a verificar se determinadas instâncias pertencem a uma classe, verificando a consistência do domínio representado. Outras linguagens com capacidade semântica para expressar conceitos e relações entre conceitos podem vir a ser usadas, como é o caso da RDFS (*Resource Description Framework Schema*) e da OWL.

A Figura 19 ilustra uma ontologia do domínio de vinhos, mostrando a sua estrutura hierárquica de categorias. A ontologia em questão foi construída na ferramenta *Protégé-2000* (HORRIDGE et al., 2004), cujo desenvolvimento foi realizado por uma comunidade médica na Universidade de *Stanford* [57], no intuito de auxiliar a padronização do vocabulário inerente ao domínio da medicina. A ferramenta permite categorizar conceitos em classes, definir propriedades desses conceitos e adicionar restrições (de quantificador, de cardinalidade e de

---

[56] "[...] the term ontology [...] classification of basic categories."
[57] http://protege.stanford.edu

valor, os quais restringem as instâncias que pertencem a uma classe). Possui suporte a vários formatos de representação como *XML Schema, RDF e OWL*. Além do *Protégé*, outras ferramentas para edição de ontologias podem ser encontradas em Almeida e Bax (2003, pg.19).

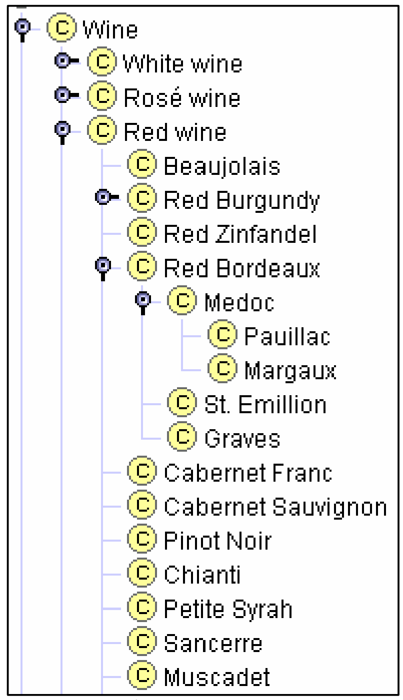


Figura 19 – Uma ontologia do domínio de vinhos desenvolvida no *Protégé 2000*
Fonte: Adaptado de Noy e McGuinness (2001)

Apesar da falta de consenso entre alguns autores em relação à classificação de tipos de ontologias (MIZOGUCHI; VANWELKENHUYSEN; IKEDA, 1995; VAN HEIJIST; SCHREIBER; WIELINGA, 2002), Guarino (1998) classifica ontologias em quatro categorias com base em seu conteúdo: de *alto nível*, de *domínio*[58], de *tarefas* e de *aplicação*. As ontologias de *alto nível* descrevem conceitos bastante gerais, tais como, espaço, tempo, matéria, objeto, etc., que são independentes de um domínio particular. As ontologias de *domínio* descrevem um vocabulário relacionado a um domínio genérico, tal como medicina. As ontologias de *tarefas* também expressam um vocabulário inerente a um domínio genérico ou uma tarefa genérica, como vendas. Ambos os tipos são especializações de termos advindos da ontologia de alto nível. Percebe-se que a definição de ontologia de domínio e de tarefa se sobrepõe, estabelecendo certa confusão em relação ao termo a ser utilizado. E, finalmente, as ontologias de *aplicação* descrevem conceitos dependentes de domínio e de tarefa particulares, que geralmente são especializações dos termos advindos das ontologias de domínio e de

[58] Ressalta-se aqui o tipo de ontologia que se pretende construir a partir da proposta metodológica enfatizada neste estudo: *ontologias de domínio*.

tarefa. A Figura 20 apresenta os níveis de dependência entre os tipos de ontologias e os relacionamentos de especializações existentes entre elas.

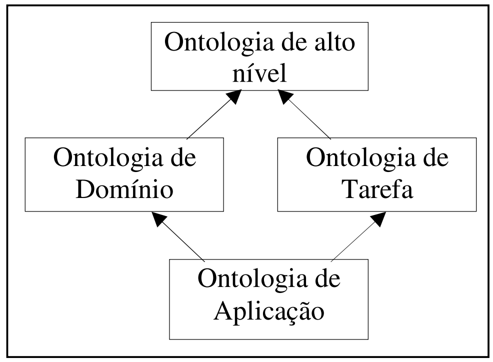


Figura 20 – Tipos de Ontologias
Fonte: Adaptado de Guarino (1998)

Chandrasekaran, Johnson e Benjamins (1999) citam alguns exemplos de ontologias de alto nível, como a Cyc (LENAT, 1995), a Wordnet (MILLER, 1995) e a de Sowa (SOWA, 1999; SOWA, 2000). E comentam sobre problemas existentes na organização de tais ontologias. O problema está na análise e definição dos conceitos mais gerais da ontologia, resultando em divergências terminológicas entre ontologias nesse nível, como acontece, por exemplo, com a categoria *thing* da ontologia Cyc, que possui as subcategorias *individual object*, *intangible* e *represented thing*. Já na ontologia Wordnet a categoria *thing* possui as subcategorias *living thing* e *nonliving thing*. Em síntese, foram construídas a partir de um mecanismo de abstração diferente.

Os autores assinalam que as diferenças aparecem porque algumas dessas ontologias, em princípio, possuem diferenças taxonômicas, além de algumas delas não servirem de propósito ferramental para construção de ontologias, ou mesmo não serem consideradas explicitamente como ontologias. Entretanto, os autores advogam que mesmo com as diferenças aparentes, há um entendimento comum sobre ontologias, vistas como representação de coisas: a) existem objetos no mundo; b) objetos possuem propriedades ou atributos que podem assumir valores; c) objetos podem fazer relações com outros objetos; d) propriedades e relações podem mudar a qualquer momento; e) existem eventos que ocorrem em diferentes momentos; f) existem processos nos quais objetos participam e ocorre o término desse processo; g) o mundo e seus objetos podem ser de uma forma a depender do estado; h) eventos podem causar outros eventos; e i) objetos podem ter partes. A representação dos fatos deverá ser feita pelo modelador do domínio, que deverá levar em consideração todas as perspectivas citadas acima.

As ontologias de *domínio* são o tipo mais comumente desenvolvido, sendo que diversos trabalhos podem ser encontrados na literatura, como na área de química (GÓMEZ-PEREZ, FERNANDEZ e VICENTE, 1996), modelagem de negócio (GRUNINGER e FOX, 1995), bioinformática (STEVENS et al., 2004) , modelagem de processos de *software* (FALBO, 1998), entre outros. Ontologias classificadas como sendo de *domínio* necessitam de ontologias de alto nível para concepção da estrutura inicial, conforme mostrado na Figura 20.

2.3.3.3 Aplicações que fazem uso

Guarino (1998) mostra o papel das ontologias nos principais componentes de um sistema de informação automatizado: recursos de informação (bancos de dados e bases de conhecimento), interface com o usuário e programas de aplicação. Como um componente de banco de dados a ontologia poderia ser útil como um artefato resultante da análise de requisitos, ou seja, uma especificação explicita dos requisitos para bancos de dados e, posteriormente, auxiliar na fase de modelagem. Além da possibilidade de integrar bancos de dados heterogêneos a partir de ontologias disponíveis para mediação no processo de integração. Como um componente de interface, o usuário poderia, por exemplo, consultar e navegar na ontologia disponível no sistema de informação a fim de recuperar as informações de forma precisa e em tempo hábil. E, como componente de aplicativo, o desenvolvedor do aplicativo poderia fazer uso da ontologia para gerar parte do sistema de informação, além do aplicativo contar com uma representação explicita do conhecimento acerca do código do programa que pode vir a ser útil na comunicação entre agentes[59] computacionais para uma dada finalidade.

Chandrasekaran, Johnson e Benjamins (1999) comentam sobre a necessidade do uso de ontologias em algumas comunidades para fins de organização das informações em seus sistemas de informações. É o caso dos sistemas de recuperação de informação, bibliotecas digitais, integração de fontes de informações heterogêneas e máquinas de busca na internet. Essas últimas vêm utilizando ontologias na construção de categorias e subcategorias a fim de realizar consultas semânticas que melhorem as clássicas buscas por palavra-chave.

As ontologias também vêm se destacando no campo de Ciência da Informação em aplicações voltadas a representação de conteúdo em ambientes virtuais e construção de interfaces cooperativas (JORGE, 2005), recuperação de informação e gestão de conhecimento

[59] Um agente computacional possui atributos que possam distingui-lo de simples programas, tais como operar sob controle autônomo, perceber seu ambiente, adaptar-se a mudanças e ser capaz de assumir metas de outros agentes (RUSSELL e NORVIG, 2004).

(ALMEIDA, 2006; CERQUEIRA, 2007) e interoperabilidade de sistemas de informações (ALMEIDA, 2002). Souza (2006, p. 169) traça um panorama atual de estratégias alternativas para melhoria dos SRIs, destacando dentre uma delas o papel das ontologias em ambientes virtuais, como por exemplo em bibliotecas digitais. Através dos metadados as ontologias são concebidas, expressando uma tentativa de um conhecimento consensual acerca das diversas fontes de informação existentes em tal ambiente. Dessa forma, as ontologias podem ampliar os pontos de acesso de documentos, através da representação mais significativa dos mesmos e permitir que as informações disponíveis sejam interpretadas pelas máquinas, possibilitando que os sistemas de informações melhorem do ponto de vista de interoperabilidade.

A Web Semântica é um exemplo de aplicação que faz uso de ontologias, e é responsável pela compatibilização de conceitos encontrados em bancos de dados dos mais diversos tipos. Como aponta Tim Berners-Lee “dois bancos de dados podem usar diferentes identificadores para o mesmo conceito”. A solução para esse problema seria prover uma ontologia, a qual ele define como “um documento ou arquivo que define formalmente os relacionamentos entre termos”. Uma ontologia para Web tipicamente possui uma taxonomia e um conjunto de regras de inferência (GILCHRIST, 2003). A Web Semântica é abordada brevemente na próxima seção no sentido de relacionar os assuntos *padronização de conceitos*, *semântica* e *compartilhamento de informações* através do uso de ontologias.

2.3.3.4 Linguagens para Representação de Ontologias

Até o momento foram apresentadas as contribuições das ontologias para a área de representação do conhecimento. Nessa seção, as ontologias são apresentadas como um modelo formal para fins de representação de conhecimento, destacando a sua principal característica de ser uma linguagem **expressiva** na representação de seus conceitos e relações.

Foram apresentados em seções anteriores alguns modelos de representação de conhecimento como as redes semânticas, os *frames*, o modelo de dados relacional e o modelo orientado a objetos. Pôde-se observar a ausência de semântica formal nesses modelos que podem restringir o uso dos mesmos quando da necessidade de representação de domínios rigorosos como, por exemplo, o domínio da medicina.

As pesquisas na área de Engenharia de ontologias buscam resolver este problema através do estabelecimento de padrões de modelos, taxonomias, vocabulários e terminologias de forma a se conseguir clareza, coerência e comprometimento ontológico sobre um domínio de conhecimento (GRUBER, 1993; GUARINO, 1998), conforme já foi discutido

anteriormente. Contudo cabe aqui destacar novamente tais características especiais. No que diz respeito à clareza, uma ontologia deveria ser capaz de expressar claramente o significado de cada termo; ser coerente com as demais definições, isto é, uma sentença inferida não poderia contradizer outra definição; e exigir comprometimento ontológico para fins de compartilhamento de conhecimento, permitindo que a comunidade de usuários envolvida possa expressar um conhecimento consensual conforme necessário. Para tal, uma linguagem capaz de fornecer subsídios para representar um modelo com tais características é fundamental na implementação da ontologia.

Os metadados são a forma mais comumente empregada para acrescentar semântica a informações (GILLILAND-SWETLAND, 2000). Um metadado pode ser representado de diversas formas, mas a maneira mais comum encontrada atualmente é através das chamadas **linguagens de marcação** (do inglês *markup languages*). Estas linguagens, cujo padrão mais conhecido e utilizado é o XML, definem *tags* ou marcações que são adicionadas aos dados a fim de indicar alguma informação importante.

No campo das ontologias, várias linguagens baseadas em XML têm sido propostas para representar ontologias. Por exemplo, RDF e RDF *Schema* (LASSILA e SWICK, 1999), XOL – *XML Based Ontology Exchange Language* (KARP, CHAUDHRI e THOMERE, 1999), SHOE – *Simple HTML Ontology Extension* (HEFLIN e HENLER, 2000), OIL (FENSEL et al., 2001), DAML+OIL (HORROCKS et al., 2001) e OWL (DEAN et al., 2003). Tais linguagens, também chamadas de "linguagens baseadas na Web", encontram-se ainda em fase de desenvolvimento e em constante evolução (LOZANO-TELLO e GOMEZ-PEREZ, 2004). Ainda que o padrão XML tenha se tornado bastante popular, logo se percebeu que somente esse padrão não é suficiente para permitir a correta interpretação das informações por um sistema informatizado, pois tal sistema não consegue inferir, através das marcações, o que uma informação significa. Várias alternativas foram e ainda estão sendo propostas para este problema pelo *World Wide Web Consortium* (W3C) [60] como: possibilitar adicionar metadados sobre os metadados (informação sobre as marcações utilizadas); padronizar os nomes de marcações; especificar a maneira como a informação deveria ser representada (eliminado possíveis inconsistências); determinar um identificador único para cada marcação de forma a evitar marcações homônimas; descrever metadados mais complexos, por exemplo, como representar que um determinado conceito é subclasse de outro conceito, ou que uma determinada propriedade é transitiva, ou qual a cardinalidade máxima e mínima de uma

[60] http://www.w3.org

propriedade. Todos esses problemas passaram a ser objeto de estudos com o surgimento da Web Semântica.

O principal objetivo da Web Semântica é criar padrões e tecnologias que permitam acrescentar semântica a informações de forma que um sistema informatizado seja capaz de entender o que a informação significa (BERNERS-LEE, HENDLER e LASSILA, 2001). Isto pode vir a ser útil nos mecanismos de busca existentes hoje na Internet. A semântica envolvida nas páginas Web permitirá uma busca mais precisa nas informações disponíveis (NARDON, 2003).

Tim Berners-Lee, idealizador e principal arquiteto dessa iniciativa, afirma que a Web Semântica é criada a partir de um conjunto de especificações, conforme mostra a Figura 21 (BERNERS-LEE, 2003).

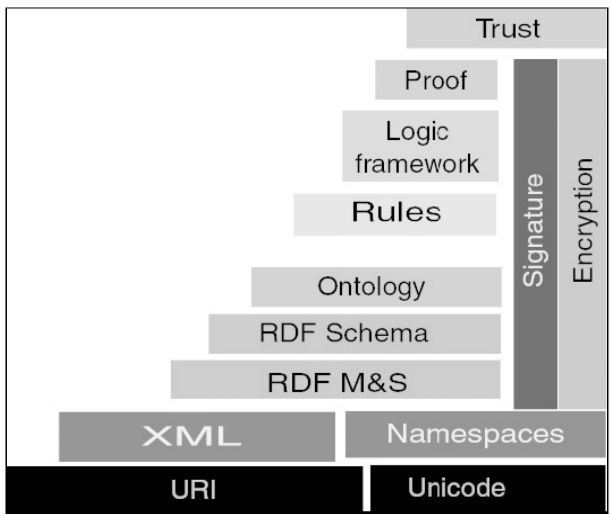


Figura 21 - Conjunto de especificações para a Web Semântica.
Fonte: adaptado de Breitman (2005, p. 49).

Para entender a camada da Web Semântica chamada de *Ontology* (ontologia) é necessário analisar as camadas anteriores propostas pelo W3C, que podem ser organizadas na seqüência: i) o Identificador Uniforme de Recurso (URI) [61] é uma forma de identificar unicamente recursos, os quais podem ser itens como páginas Web, uma coluna de uma tabela, uma imagem, um documento, entre outros. Uma URL (Localizador de Recurso Universal) [62], por exemplo, é um tipo de URI usado para identificar páginas Web. O padrão Unicode[63] define um único número para cada caractere representado, não importando em que plataforma ou linguagem, sendo, portanto, fundamental para manter a interoperabilidade; ii) camada contendo padrões de metadados, XML, *namespaces* e DTD ou XML *Schema*; iii) camada

---

[61] *Uniform Resource Identifier em inglês*
[62] *Universal Resource Locator em inglês.*
[63] http://www.unicode.org

com padrão RDF; e iv) camada com padrão RDF-*Schema*[64]. Na camada *ontology* são definidos padrões para descrever ontologias, incluindo construtores que permitem especificar, por exemplo, se uma propriedade é transitiva ou não, qual a cardinalidade da mesma e quais as restrições existentes sobre as classes. Atualmente, o padrão mais utilizado nesse nível é a OWL para representação de ontologias.

Os níveis mais altos como regras, lógica e prova (*Rules*, *Logic Framework*, *Proof*) ainda encontram-se em definição e a intenção das pesquisas é permitir uma confiabilidade (*Trust*) no conhecimento obtido, sendo que cada aplicação deve avaliar se um conhecimento é confiável ou não a partir do seu contexto (BREITMAN, 2005).

Alguns termos mencionados no conjunto de especificações da Web Semântica (como *namespace,* metadados, DTD) serão elucidados nas seções seguintes quando da apresentação do histórico e evolução das linguagens supracitadas (XML, RDF, RDFS e OWL).

2.3.3.4.1 Linguagens de marcação e padrões de metadados eletrônicos

Historicamente a palavra "marcação" descreve anotações ou marcas que informavam a maneira como parte de um texto deveria ser representada. O paradigma das linguagens de marcação vislumbra que um documento é composto de conteúdo, estrutura e estilo. Segundo Bax (2001), o conteúdo seria a informação em si; a estrutura seria como se dá a organização da informação; e o estilo, a maneira como a informação será apresentada para o usuário.

Algumas linguagens de marcação de padrão aberto (SGML, HTML, XML, etc) permitem a criação de documentos que podem ser manipulados independentemente de plataforma de *hardware* e *software*. As linguagens XML (*eXtensible Markup Language*) e HTML (*HyperText Markup Language*) são linguagens com tais objetivos, sendo originadas da SGML[65] (*Standard Generalized Markup Language*). O CERN – *European Organization for Nuclear Research* (Centro Europeu de Pesquisas Nucleares) utilizou no início da década de 80 a SGML em pesquisas de hipertextos. Dessa maneira, o pesquisador Tim Berners-Lee acabou criando no início da década de 90 o *Word-Wide Web*, que empregava a idéia de ligações entre documentos (páginas) localizados em qualquer parte do mundo, através da rede mundial de

---

[64] As linguagens RDF e RDF-*Schema* podem ser consideradas a fundação da Web Semântica. Tais linguagens tornaram-se um padrão recomendado pela W3C que define uma linguagem para descrever recursos, tais como as páginas na Web. Uma recomendação do W3C é entendida pela indústria e pela comunidade como um padrão para a Internet (BREITMAN, 2005).

[65] A SGML foi reconhecida como um padrão ISO (8879) em 1986. É considerada uma linguagem para definir outras linguagens, na qual cada documento SGML carrega consigo sua própria especificação formal, o *Data Type Document* (DTD).

computadores (a Internet) e do URL. Daí o surgimento da HTML[66], sendo formada por um conjunto de marcações (conhecidas por *tags*) predefinidas, com a função de apresentar a informação por meio de páginas Web. Tal linguagem é um padrão usado em diversas aplicações, incluindo navegadores, editores, servidores de base de dados e outros. A Figura 22 apresenta um exemplo de um documento HTML simples, contendo algumas formatações para títulos e listas.

```
<HTML>
<HEAD>
   <meta name="Autor" content="João da Silva">
</HEAD>
<BODY>
<H1> Título 1 </H1>
<H2> Título 2 </H2>
<H3> Título 3 </H3>
<P> Texto Normal </P>
<LI> item 1 </LI>
<LI> item 2 </LI>
<LI> item 3 </LI>
</BODY>
</HTML>
```

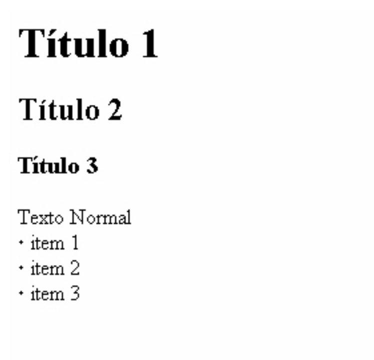


Figura 22 - Exemplo de um documento simples em HTML e sua apresentação (à direita)
Fonte: o próprio autor

Entretanto, tal padrão apenas descreve como uma página Web deve ser exibida, não oferecendo nenhuma descrição dos dados. Isso pode acarretar restrições nas buscas, no intercâmbio de dados entre plataformas diferentes, na personalização da informação, dentre outras.

A linguagem XML[67] surge então em 1997 com a proposta de ser um formato para descrição de dados semi-estruturados. Uma diferença em destaque entre HTML e XML é que esta última permite um número ilimitado de marcações, indicando o que cada elemento significa, e não o que será apresentado. Tal formato é especificado por definições de tipos de documentos (DTDs) ou por esquemas XML (XML *Schema*[68] ou XMLS). Um DTD define a estrutura e sintaxe de um documento, ajudando a validar se o mesmo está em conformidade com uma estrutura. Esquemas XML têm a mesma função, mas permitem um maior tratamento às informações. Pode-se definir tipo e formato exato dos atributos, número exato de instâncias de um aninhamento e há mecanismos de inclusão e derivação que proporcionam o seu reuso (TOLENTINO, 2004). Estas capacidades ajudam a reduzir a distância entre DTDs e

[66] http://www.w3.org/MarkUp/
[67] http://www.w3.org/XML/
[68] http:///www.w3.org/XML/Schema

ontologias, pois podem ser introduzidos meios para restringir estruturas de documentos e conteúdos, herança para elementos e atributos, tipos primitivos e outras características úteis (FENSEL et al., 2001).

A XML pode utilizar padrões de metadados como o Dublin Core[69] na declaração de vocabulários, além de possibilitar criar vocabulários pelos próprios autores do documento. Tais vocabulários são declarados através da utilização da marcação *namespace*[70], garantindo que para cada termo utilizado na confecção do documento tenha-se apenas uma definição, evitando assim ambigüidade em sua referência. Um *namespace* XML é uma coleção de nomes (usados em documentos XML como nomes de elementos e de atributos) identificados através da referência a um URI. No exemplo da Figura 23, os elementos **xmlns:dc** e **xmlns:bib** indicam respectivamente o *namespace* do próprio Dublin Core e o relativo ao domínio de biblioteca. Cada elemento está localizado dentro de um único arquivo, que contém a especificação do vocabulário utilizado para descrever a entidade livro no domínio de biblioteca. Para evitar a repetição do endereço do vocabulário todas as vezes que se fizer menção a um de seus termos, cria-se um *namespace* (mapeado a um URI) para ele. No exemplo, as etiquetas do tipo **bib:** e **dc:** substituem respectivamente a menção aos seus endereços completos (http://biblioteca/vocabulario/livro e http://purl.org/dc/elements/1.1/). Exemplos são *title*, *description*, *publisher* e *identifier* relacionados ao prefixo **dc:** ; e *preço* e *instituição* relacionados ao prefixo **bib:**.

```
<?xml version="1.0"?>
<xmlns:dc="http://purl.org/dc/elements/1.1/">
<xmlns:bib="http://biblioteca/vocabulario/livro">
<bib:livro>
    <dc:title>  Web Semântica – a Internet do Futuro </dc:title>
    <dc:description> Este livro oferece um tratamento extenso e compreensivo da Web     Semântica e das
novas tecnologias necessárias para compreender e implementá-la.
    </dc:description>
    <dc:publisher> Livros Técnicos e Científicos </dc:publisher>
    <dc:identifier>  http://www.inf.puc-rio.br/index.html </dc:identifier>
    <bib :preco> R$ 33,71 </bib :preco>
    <bib:instituicao> Pontifícia Universidade Católica do Rio de Janeiro </bib:instituicao>
</bib:livro>
```

Figura 23 – Exemplo de um documento em XML
Fonte: o próprio autor

---

[69] Informações adicionais em http: //dublincore.org/
[70] http://www.w3.org/TR/REC-xml-names/

O uso de metadados auxilia sobremaneira a indexação de páginas na Web, melhorando a eficiência na recuperação da informação. Gilliland-Swetland (2000) dá uma definição panorâmica sobre metadado: *"a soma total do que pode ser dito sobre algum objeto informacional em algum nível de agregação"*. A autora entende objeto informacional como um item ou grupo de itens que pode ser manipulado ou endereçado, independente de tipo ou formato, como um objeto singular ou discreto por um humano ou computador.

Velluci (1998) considera que o advento da Internet potencializou o esforço de bibliotecas e especialistas em informação para melhorar métodos de descrição, organização e recuperação de objetos digitalizados acessados remotamente. Este esforço envolve ainda criadores, provedores e usuários de recursos eletrônicos dos setores acadêmicos, públicos e comerciais. Alvarenga (2003) contribui nessa perspectiva dizendo que no novo contexto de produção, organização e recuperação de objetos digitais, as metas de trabalho não se restringem à criação de representações simbólicas dos documentos constantes de um acervo, mas compreendem a criação de novas formas de escrita para os hipertextos e a criação dos denominados metadados, muitos dos quais podem ser extraídos diretamente dos próprios objetos, constituindo-se, esses, em chaves de acesso a serviço dos *internautas*.

Vellucci (1998) assinala, no entanto, a importância da convergência de esforços, no intuito de conceber uma estrutura flexível para a organização e acesso a essas informações. Os padrões de metadados têm a capacidade de prover um vocabulário comum para descrever uma variedade de estruturas de dados capazes de satisfazer uma enorme gama de interessados - incluindo os cientistas da computação e engenheiros, os quais desenvolvem máquinas de busca e criam padrões de documentação para Internet, além de estudantes, bibliotecários e arquivistas que organizam e provêem acesso a recursos eletrônicos.

Diante a tais necessidades, principalmente em função da organização da Web, surgiram novas propostas capazes de tratar a semântica envolvida nas estruturas de metadados existentes nos diversos ambientes de informação. Trata-se de aplicações baseadas em XML com funções de publicar vocabulários legíveis para humanos e máquinas, além de proporcionar a reutilização e a distribuição dos metadados. Tais linguagens são apresentadas a seguir.

2.3.3.4.2 RDF – *Resource Description Framework*

Segundo o W3C, a linguagem RDF[71] é um padrão para representar metadados de recursos Web, tais como título, data de criação e direitos autorais a respeito de documentos. A linguagem pode ser usada para representar informação sobre os recursos que podem ser identificados na Web como, por exemplo, em aplicações de comércio eletrônico na disponibilização de informações sobre os itens comercializados.

A linguagem adota a sintaxe XML (também chamada de XML/RDF) e sua especificação pode ser considerada um vocabulário legível para as máquinas, que possibilita a troca de informações entre máquinas que utilizam aplicativos ou até mesmo sistemas operacionais diferentes (BREITMAN, 2005).

As unidades de informação em RDF são descritas em triplas, ou seja, no formato de **recurso**, **propriedade** e **valor**, similar a estrutura de uma rede semântica, conforme foi visto na seção 2.3.1.1. A linguagem utiliza um URI para identificar um recurso Web e uma propriedade para descrever tal recurso. O URI possibilita fazer referência não ambígua aos termos envolvidos no vocabulário utilizado pelo documento através da marcação de *namespace* provida pelo XML, conforme explanado na seção anterior.

As propriedades associadas aos recursos denotam os relacionamentos dos valores associados aos recursos. Tais valores podem ser atômicos (*strings*, números, etc) ou outros recursos, expressando relacionamentos entre os próprios recursos. A Figura 24 ilustra um exemplo de um modelo de dados representado em RDF, no qual os recursos são descritos através de seus valores. No exemplo, o *recurso web 1* possui um relacionamento com o *recurso web 2*, que por vez possui relacionamentos com valores do tipo *string*.

---

[71] http://www.w3c.org/RDF

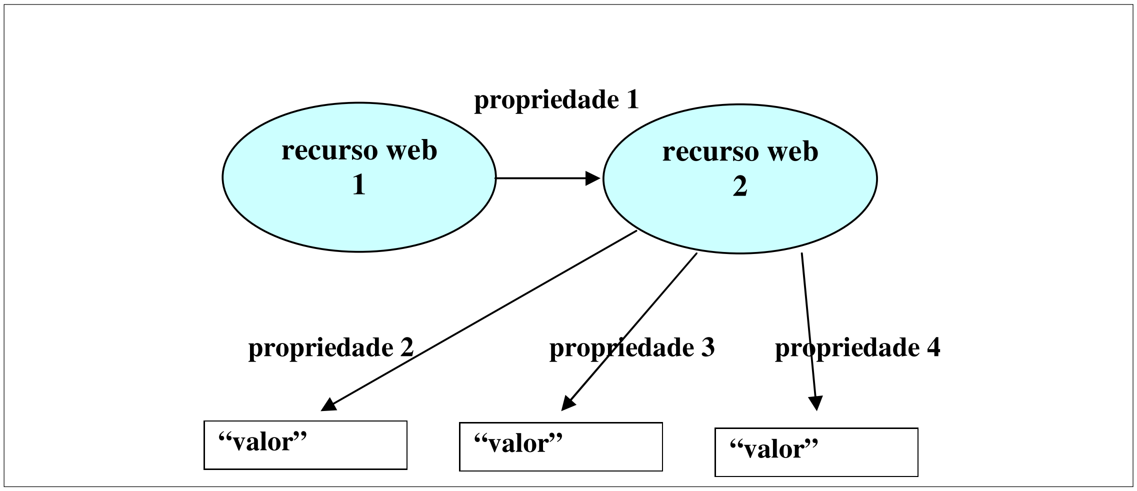


Figura 24 – Modelo de dados representado em RDF
Fonte: o próprio autor

Um exemplo de código RDF representando o modelo da Figura 24 é mostrado abaixo:

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.inf.puc-rio.br/~Karin/vocabulario/loja_de_disco/#">
<rdf:Description
rdf:about="http://www.submarino.com.br/cds_top10.asp">
  <cd:titulo> <rdf :Description
rdf:about="http://www.submarino.com.br/#Acustico MTV"></rdf :Description></cd :titulo>
</rdf :Description>
<rdf :Description
rdf:about="http://www.submarino.com.br/#Acustico MTV">
  <cd:artista>Marcelo D2 </cd:artista>
  <cd:gravadora>Sony </cd:gravadora>
  <cd:pais>Brasil </cd:pais>
</rdf :Description>
</rdf :RDF>
```

Figura 25 – Exemplo de código em RDF
Fonte: adaptado de Breitman (2005)

O elemento **rdf:RDF** (que tem o **/rdf:RDF** como fechamento da marcação) indica que o conteúdo é RDF. Logo em seguida aparecem os *namespaces* indicando a especificação do vocabulário de RDF e o vocabulário relativo ao domínio da loja de disco. O elemento **rdf:Description** (e seu fechamento **/rdf:Description**) representa os recursos *recurso web 1* e *recurso web 2* através de seus URIs identificados pelo atributo **rdf:about**. A propriedade do *recurso web 1* é representada pelo elemento **cd:titulo**, que estabelece a relação com o *recurso*

web 2, que por sua vez possui três propriedades representadas pelos elementos **cd:artista**, **cd:gravadora** e **cd:pais** com seus respectivos valores “Marcelo D2”, “Sony” e “Brasil”.

O RDF fornece uma semântica simplificada que pode ser representada através da sintaxe XML. Contudo, um esquema seria necessário para organizar e melhor especificar os recursos RDF. Surge assim a linguagem RDF-*Schema* descrita a seguir.

2.3.3.4.3 RDFS – ***Resource Description Framework Schema***

A RDF se restringe em criar instâncias e associar propriedades descritivas a cada uma delas, sem uma estrutura de dados que possa ser estendida através do refinamento de subclasses (BRAGANHOLO e HEUSER, 2001). A RDF-*Schema*[72] foi concebida para impor uma estrutura ao RDF no sentido de possibilitar ao usuário definir os dados no formato de um esquema, isto é, em classes, subclasses, propriedades, domínios e restrições.

O W3C define a RDF-*Schema* como “[…] um padrão que descreve como usar RDF para descrever vocabulários RDF na Web” [73]. A RDF permite asserção de simples sentenças consistindo um sujeito, um predicado e um objeto. A linguagem não descreve qual o significado do sujeito, do predicado e do objeto, além de não descrever o significado dos relacionamentos entre eles. Já a RDFS fornece as primitivas básicas para a criação de ontologias simples, incluindo relacionamentos de generalização para classes e propriedades, além da definição de como recursos podem ser descritos como sendo membros de uma ou mais classes (MCBRIDE, 2004; BREITMAN, 2005).

Uma classe representa uma coleção de recursos e, geralmente, é identificada por um URI. Um recurso (descrito no documento RDF) pode ser declarado para ser membro de uma classe, isto é, o recurso é declarado como instância dessa classe. Para definição de uma classe e do relacionamento de herança entre duas classes utiliza-se os elementos **rdfs:Class** e **rdfs:subClassOf**, conforme ilustrado no código da Figura 26.

---

[72] http://www.w3c.org/TR/rdf-schema/

[73] “[…] a standard which describes how to use RDF to describe RDF vocabularies on the Web”

```
<rdf:RDF xml:base="http://example.org/example">
  <rdfs:Class rdf:ID="Person"/>
  <rdfs:Class rdf:ID="Man">
     <rdfs:subClassOf rdf:resource="#Person"/>
  </rdf:Class>
</rdf:RDF>
```

Figura 26- Um exemplo de herança entre duas classes
Fonte: Mcbride (2004)

O exemplo da Figura 26 representa que a classe *man* (homem) é uma subclasse da classe *person* (pessoa). Isso implica que os membros da classe *man* herdam as propriedades da classe *person*.

As propriedades em RDF também formam uma classe. O RDFS define **rdf:Property** como sendo a classe de todas as propriedades RDF. Formalmente, uma propriedade é um recurso que tem uma extensão de propriedade que é um conjunto de pares de objetos relacionados através de uma propriedade. Uma propriedade pode ser uma *subpropriedade* da outra, que significa que qualquer um dos objetos relacionados através da *subpropriedade* são também relações da *superpropriedade*. O exemplo da Figura 27 mostra a propriedade *husbandOf* (marido de) sendo definida como *subpropriedade* da propriedade *partnerOf* (cônjuge de).

```
<rdf:RDF xml:base="http://example.org/example">
  <rdf:Property rdf:ID="partnerOf/">
  <rdf:Property rdf:ID="husbandOf">
    <rdfs:subPropertyOf rdf:resource="#partnerOf/">
  </rdf:Property>
</rdf:RDF>
```

Figura 27 - Um exemplo de herança entre propriedades
Fonte: Mcbride (2004)

Algumas restrições podem ser impostas aos vocabulários. O RDFS fornece as primitivas **rdfs:domain** que especifica o domínio de uma propriedade, ou seja, a classe de recursos que pode aparecer como sujeito da sentença RDF; e **rdfs:range** que especifica o alcance da propriedade, ou seja, a classe de recursos que pode aparecer como objeto da sentença RDF.

A propriedade **rdfs:domain** é usada para declarar que os sujeitos associados a uma propriedade são instâncias de uma ou mais classes. Se uma propriedade tem mais de uma classe em seu domínio, significa que o sujeito associado àquela classe deve ser instância de

todas as classes que compõem o domínio. Já a propriedade **rdfs:range** é usada para declarar que os objetos, ou seja, os valores de uma propriedade, são instâncias de uma ou mais classes. Estas classes formam o escopo da propriedade. Se uma propriedade tem mais de uma classe em seu escopo, significa que seus valores devem ser instâncias de todas as classes que compõem o escopo.

O código apresentado na Figura 28 impõe a restrição de que o sujeito da sentença RDF (instância João) será sempre uma instância das classes *Man* e *MarriedPerson*. E, o valor do objeto (instância Maria) será sempre instância das classes *Woman* e *MarriedPerson*.

```
<rdf: Property rdf:ID="husbandOf">
  <rdfs:domain rdf:resource="#Man"/>
  <rdfs:domain rdf:resource="#MarriedPerson"/>
   <rdfs:range rdf:resource="#Woman"/>
   <rdfs:range rdf:resource="#MarriedPerson"/>
</rdf:RDF>
```

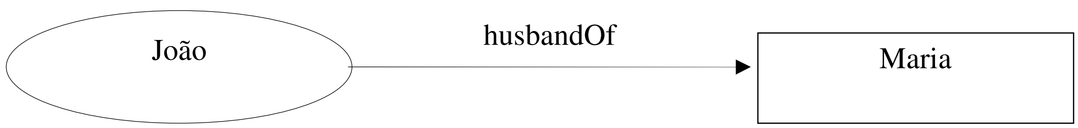


Figura 28 - Um exemplo de restrições no vocabulário RDF
Fonte: adaptado de Mcbride (2004)

O RDFS pode ser considerado um tipo de dicionário que pode ser lido por máquinas (BREITMAN, 2005). Conforme foi visto, a linguagem fornece um conjunto de primitivas que permitem a modelagem de ontologias simples, por exemplo, através do uso de *subClassOf* e *subPropertyOf* na definição de uma taxonomia, além de possibilitar as restrições no vocabulário através da definição de domínio e de escopo (*domain* e *range*) nas propriedades dos recursos RDF. No entanto, RDFS tem sido criticada como linguagem para ontologias pela falta de expressividade de seus construtos, principalmente pela inexistência de relação de equivalência, relação inversa e restrições de cardinalidade encontradas em outras linguagens para desenvolvimento de ontologias (MCBRIDE, 2004). Tal deficiência pode ser suprida através de sua extensão encontrada na linguagem OWL apresentada a seguir.

2.3.3.4.4 OWL e Ontologias

A OWL (*Web Ontology Language*, ou linguagem de ontologias para a Web) é a linguagem desenvolvida para representar ontologias, criada pelo *Web Ontology Working*

Group, da W3C. A OWL deriva de um consenso entre a proposta européia OIL[74] (*Ontology Inference Layer*, camada de inferência para ontologias, ou ainda, *Ontology Interchange Language*, linguagem de intercâmbio em ontologias) e a DAML (DARPA *Agent Markup Language*, linguagem de anotação para agentes do Departamento de Defesa dos Estados Unidos). Similar à DAML+OIL[75], a intenção da OWL é representar conceitos e seus relacionamentos na forma de uma ontologia.

De acordo com o W3C, "a linguagem OWL é projetada para ser usada por aplicações que necessitam de processar o conteúdo das informações, ao invés de apenas apresentar informações aos seres humanos. Além de facilitar a interpretação por máquinas de conteúdos na Web em XML, RDF e RDFS através de um vocabulário e uma semântica formal "[76].

Os requisitos de uma boa linguagem para a criação e representação de uma ontologia são: a) ter uma sintaxe bem definida; e b) ter uma semântica formal para que máquinas possam realizar inferências. Antoniou (2004, p.110) afirma que ter uma sintaxe bem definida é condição necessária para o processamento de informação por máquinas. As linguagens DAML + OIL e OWL foram construídas em cima de padrões bem aceitos na Web como RDF e RDFS. O conceito de semântica formal é pré-requisito para suportar inferências (*reasoning*) e checar a consistência da ontologia. A semântica "escreve" o significado do conhecimento "precisamente" (não o conhecimento subjetivo), consagrado na lógica matemática, conforme foi visto na seção 2.3.1.3. Pode-se, por exemplo, realizar inferências sobre:

- Elemento de uma classe: Se x é instância da Classe C, e C é uma subclasse de D, então se pode inferir que x é instância de D.
- Equivalência de classes: Se a classe A é equivalente a Classe B, e B é equivalente à classe C, então A é equivalente a C.
- Consistência: Se x é instância da classe A e A é subclasse de B ∩ C, A é subclasse de D, e B e D são "disjuntas" conclui-se que há uma inconsistência, tendo em vista que A deveria estar vazia, entretanto tem uma instância x. É uma indicação de erro na ontologia.
- Classificação: a partir da declaração de uma condição suficiente verifica-se a instanciação de objetos nas classes.

---

[74] http://www.ontoknowledge.org/oil
[75] http://www.daml.org/2001/03/daml+oil-index.html
[76] "[…] is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics"

Diante disso, torna-se visível que a semântica formal empregada na representação de conceitos e relações da ontologia, em conjunto com um mecanismo de inferência, possibilitaria que várias verificações de casos fossem feitas na ontologia. Tal suporte seria uma das vantagens das ontologias para com os vocabulários controlados tradicionais da biblioteconomia, como os tesauros. O instrumento tesauro, por exemplo, num contexto computacional poderia contar com um mecanismo de inferência para classificar automaticamente um termo em sua estrutura conceitual, facilitando, desta forma, o trabalho do profissional indexador quando do surgimento de um novo termo a ser classificado e direcionado ao sistema.

Segundo Antoniou (2004, p.111), a semântica formal e o suporte a inferências são geralmente fornecidos através do mapeamento da linguagem ontológica para um formalismo lógico conhecido, e através da utilização de inferências automáticas já existentes nesses formalismos. A linguagem OWL está mapeada (parcialmente) em lógica descritiva e faz uso de inferências existentes como FaCT[77] (*Fast Classification of Terminologies*) e RACER[78]. Conforme já explanado, a OWL baseia-se na linguagem RDF e RDFS, porém define classes e relacionamentos de forma mais detalhada. Uma crítica ao RDFS é a falta de semântica formal, ou seja, não há mapeamento dessa linguagem para nenhum tipo de lógica formal, o que elimina qualquer possibilidade de suporte a deduções e inferências (que são fatores cruciais para processamento automatizado). Diante a tais limitações e com a intenção de obter esforços para o desenvolvimento de uma linguagem de representação de conhecimento capaz de criar e compartilhar ontologias na web, o *Web Ontology Working Group* especificou a OWL em três diferentes *sublinguagens*, de acordo com uma ordem decrescente de expressividade: a) OWL Full; b) OWL DL e c) OWL Lite.

No caso da *OWL Full*, a linguagem é considerada completa, ou seja, usa todas as primitivas da linguagem OWL e permite a combinação arbitrária de tais primitivas com RDF e RDFS. A vantagem da OWL Full é que ela é completamente compatível com RDF sintática e semanticamente, ou seja, qualquer documento válido RDF é também um documento válido OWL Full, e qualquer conclusão válida num esquema RDF é também uma conclusão válida em OWL Full. No entanto, fazendo este uso mais complexo, não há garantia de suporte por máquinas (computacionalmente intratável). A *OWL DL* (*Description Logic*) utiliza a lógica

---

[77] HORROCKS, I. *The FaCT system*. In Harrie de Swart, editor, Proc. of the Int. Conf. on Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX-98), volume 1397 of Lecture Notes In Arti_cial Intelligence, pages 307{312. Springer-Verlag, 1998.

[78] HAARSLEV, V.; MOLLER, R. RACER *system description*. In Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR-01), volume 2083 of Lecture Notes In Arti_cial Intelligence. Springer-Verlag, 2001.

descritiva como base em seus construtos. A vantagem é que permite suporte a inferências e a desvantagem é a perda de compatibilidade com o RDF. E, finalmente, a *OWL Lite*, que abrange a expressividade dos *frames* e da lógica descritiva, porém com algumas restrições. Por exemplo, a cardinalidade máxima ou mínima assume apenas os valores 0 ou 1. A vantagem é o fácil entendimento da linguagem pelos usuários e a facilidade de implementação por parte de ferramentas como o *Protégé-owl*[79]. A desvantagem é a expressividade limitada (ANTONIOU, 2004; BREITMAM, 2005). As linguagens menos expressivas (OWL Lite e OWL DL) estão contidas dentro das mais expressivas (OWL DL e Full), de maneira que uma ontologia definida numa linguagem menos expressiva é aceita por uma linguagem mais expressiva. Desta forma, o que norteia a escolha da linguagem OWL a utilizar é a expressividade pretendida, ou seja, vai depender do grau de formalismo requerido na representação do domínio, conforme é colocado por Uschold e Gruninger (1996, p.6): o vocabulário pode ser representado com distintos graus de formalismo, ou seja, desde o altamente informal (*highly informal*), utilizando-se linguagem natural, até o rigorosamente formal (*rigorously formal*), utilizando-se declarações lógicas.

No que tange à sintaxe OWL, os elementos básicos da linguagem são: a) *header*; b) classe; c) propriedade e d) restrições de propriedades. Para a construção de tais elementos, a OWL utiliza o formalismo da lógica descritiva, conforme discutido na seção 2.3.1.3.

O elemento *header* (cabeçalho) é declarado para especificar os vocabulários utilizados na ontologia. Uma ontologia típica em OWL começa com um conjunto de declarações de *namespaces*. De acordo com a Figura 29, as classes a serem definidas estarão localizadas no *namespace* da primeira definição (**file:/G:/myclasses**). A segunda definição (**eyeglass**) serve para que ontologias externas possam referenciar a ontologia sendo definida. As declarações seguintes localizam as primitivas de OWL, RDF, RDFS e XML *Schema*, respectivamente.

```
<rdf:RDF
    xmlns="file:/G:/myclasses#"
    xmlns:eyeglass="file:/G:/Glasses#"
    xmlns:owl="http://www.w3.org/2003/02/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#">
```

Figura 29 – Declarações de *namespaces* na especificação da ontologia
Fonte: Freitas (2003)

[79] http://protege.stanford.edu/overview/protege-owl.html.

Uma vez que os *namespaces* foram definidos, é comum incluir uma coleção de sentenças sobre a ontologia agrupadas sob a marcação **owl:Ontology**. Tal marcação é responsável por registrar comentários, controle de versão e pela inclusão de conceitos e propriedades de outras ontologias. A Figura 30 mostra um exemplo da marcação **owl:Ontology**. É válido destacar aqui a marcação **owl:imports** cuja propriedade é transitiva, tendo conseqüências lógicas para a ontologia envolvida: se a ontologia *A* importar a ontologia *B*, e esta importar a ontologia *C*, então a ontologia *A* também importará a ontologia *C*.

```
<owl:Ontology rdf:about=" ">
    <rdfs:comment>An example OWL ontology</rdfs:comment>
    <owl:priorVersion
        rdf:resource="http://www.mydomain.org/uni-ns-old/">
    <owl:imports
        rdf:resource="http://www.mydomain.org/persons/">
    <rdfs:label>University Ontology</rdfs:label>
</owl:Ontology>
```

Figura 30 – Exemplo de informações adicionais sobre uma ontologia
Fonte: Antoniou (2004, p.116)

O elemento **classe** representa um conjunto ou coleção de indivíduos (objetos, pessoas, coisas) que compartilham de um grupo de características que os distinguem dos demais (BREITMAN, 2005). A declaração de uma classe em OWL se dá pelo elemento **<owl:Class rdf:ID="NOMEDACLASSE">.** As classes podem ser construídas de várias formas: por herança, união, interseção, complemento, pela enumeração de instâncias ou por restrições de propriedades.

O exemplo apresentado na Figura 31 mostra um cenário de especialização da classe *associateProfessor* (professor associado) para com a superclasse *academicStaffMember* (membro do corpo docente).

```
<owl:Class rdf:ID="associateProfessor ">
    <rdfs:subClassOf rdf:resource="#academicStaffMember"/>
</owl:Class>
```

Figura 31 – Definição de classe e especialização em OWL
Fonte: Antoniou (2004, p.117)

No exemplo da Figura 32, a classe *associateProfessor* é disjunta das classes *professor* (professor) e *assistantProfessor* (professor assistente) utilizando-se dos elementos **owl:disjointWith**.

```
<owl:Class rdf:about="#associateProfessor ">
    <owl:disjointWith rdf:resource="#professor"/>
    <owl:disjointWith rdf:resource="#assistantProfessor"/>
</owl:Class>
```

Figura 32 – Exemplo de formalização de uma classe em OWL
Fonte: Antoniou (2004, p.117)

No que diz respeito ao elemento **propriedade**, existem dois tipos para descrever os predicados: *propriedades de objeto* e *propriedades de tipos de dado*. O primeiro tipo relaciona objetos com outros objetos e o segundo relaciona objetos com seus valores de tipos de dado. Este último utiliza o *datatypes* do XML *Schema*, no qual faz uso da arquitetura de camada da Web Semântica, conforme foi mostrado na seção 2.3.3.4. O exemplo da Figura 33 relaciona todos os seres vivos com a propriedade ano de nascimento, cujo tipo de dado é um inteiro positivo.

```
<owl:DatatypeProperty rdf:ID="AnoNascimento">
    <rdfs:domain rdf:resource="#SerVivo" />
    <rdfs:range rdf:resource="&xsd:positiveInteger" />
</owl:DatatypeProperty>
```

Figura 33 – Exemplo de propriedade datatype
Fonte: Breitman (2005, p. 63)

As Figuras 34 e 35 exemplificam o tipo de propriedade de objetos. O primeiro exemplo além de definir o domínio e o escopo da propriedade *teaches* (ensina), define o seu relacionamento inverso através do elemento **owl:inverseOf**. Por exemplo: se Pedro *ensina* a disciplina matemática, a disciplina matemática é *ensinada* (*isTaughtBy*) por Pedro.

```
<owl:ObjectProperty rdf:ID="teaches">
    <rdfs:range rdf:resource="#course"/>
    <rdfs:domain rdf:resource="#academicStaffMember"/>
    <owl:inverseOf rdf:resource="#isTaughtBy"/>
</owl:ObjectProperty >
```

Figura 34 – Exemplo de propriedade de objetos
Fonte: Antoniou (2004, p.118)

O segundo exemplo mostra a utilização da equivalência de propriedades através do uso do elemento **owl:equivalentProperty**. Por exemplo: a propriedade *lecturesIn* (leciona em) equivale a propriedade *teaches* (ensina).

```
<owl:ObjectProperty rdf:ID="lecturesIn">
   <owl:equivalentProperty rdf:resource="#Teaches"/>
</owl:ObjectProperty >
```

Figura 35 – Exemplo de propriedade de objetos
Fonte: Antoniou (2004, p.119)

Em OWL, **propriedades** são utilizadas para criar **restrições,** representadas pelo elemento **owl:Restriction**. Como o próprio nome sugere, uma restrição é utilizada para definir alguns limites para indivíduos que pertencem a uma classe. As restrições podem ser de três tipos: a) restrições que utilizam quantificadores; b) restrições do tipo *hasValue* (tem valor de) e c) restrições de cardinalidade.

As restrições que utilizam quantificadores podem vir a ser de dois tipos: a) quantificador existencial: indica a existência de **pelo menos um** indivíduo, e em OWL é representado pela expressão *someValuesFrom*; e b) quantificador universal: pode ser interpretado como **todos** os indivíduos e é representado em OWL pela expressão *allValuesFrom.* No exemplo da Figura 36, o elemento **owl:allValuesFrom** é usado para especificar a classe de possíveis valores que a propriedade especificada pelo elemento **owl:onProperty** pode assumir: apenas indivíduos da classe *professores universitários* (*professors*) são permitidos como valores da propriedade *é ensinado por* (*isTaughtBy*) que liga à classe *primeiro ano de curso* (*firstYearCourse*). Em termos lógicos significa dizer que **todos** os indivíduos que ensinam uma instância da classe *primeiro ano de curso* devem ser *professores universitários*. Portanto, têm-se uma restrição de quantificador universal.

```
<owl:Class rdf:about="#firstYearCourse">
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#isTaughtBy"/>
         <owl:allValuesFrom rdf:resource="#Professor"/>
      </owl:Restriction>
   </rdf:subClassOf>
</owl:Class >
```

Figura 36 – Exemplo de utilização do quantificador universal
Fonte: Antoniou (2004, p.119)

No exemplo da Figura 37, o elemento **owl:someValuesFrom** é usado para declarar que todos os *membro do corpo docente* (*academicStaffMember*) devem ensinar (*teaches*) **ao menos um** *curso de graduação* (*undergraduateCourse*). Em termos lógicos significa dizer que

existe um *curso de graduação* ensinado por uma instância da classe *membro do corpo docente*, tendo-se, portanto uma restrição de quantificador existencial.

```
<owl:Class rdf:about="#academicStaffMember">
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#teaches"/>
         <owl:someValuesFrom rdf:resource="#undergraduateCourse"/>
      </owl:Restriction>
   </rdf:subClassOf>
</owl:Class >
```

Figura 37 – Exemplo de utilização do quantificador existencial
Fonte: Antoniou (2004, p.120)

O exemplo da Figura 38 mostra a determinação de um valor especifico que uma propriedade deve assumir: os indivíduos pertencentes à classe *curso de matemática* (*mathCourse*) são *lecionados* (*isTaughtBy*) pelo *professor* David Billington (representado pela identificação “949352”).

```
<owl:Class rdf:about="#mathCourse">
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#isTaughtBy"/>
         <owl:hasValue rdf:resource="#949352"/>
      </owl:Restriction>
   </rdf:subClassOf>
</owl:Class >
```

Figura 38 – Exemplo de utilização da restrição owl:hasValue
Fonte: Antoniou (2004, p.120)

E, finalmente, o exemplo da Figura 39 mostra uma situação de restrição de cadinalidade[80] mínima e máxima: um elemento pertencente à classe departamento (#*department*) deve ter no mínimo 10 membros e no máximo 30 membros, cujos literais “10” e “30” são interpretados como inteiros não negativos através da declaração de *namespace* (feita no elemento *header*) que referencia o documento XML *Schema*.

[80] É possível também especificar um número preciso, por exemplo, um estudante deve ter exatamente 2 orientadores. Isso pode ser feito usando o mesmo número em owl:minCardinality e owl:maxCardinality. Por conveniência, Owl oferece também o elemento owl:cardinality. É válido ressaltar que na owl lite a cardinalidade máxima ou mínima assume apenas os valores 0 ou 1 (ANTONIOU, 2004) .

```
<owl:Class rdf:about="#department">
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#hasMember"/>
            <owl:minCardinality
                  rdf:datatype="&xsd;nonNegativeInteger">
               10
            </owl:minCardinality>
      </owl:Restriction>
   </rdf:subClassOf>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#hasMember"/>
            <owl:maxCardinality
                  rdf:datatype="&xsd;nonNegativeInteger">
               30
            </owl:maxCardinality>
      </owl:Restriction>
   </rdf:subClassOf>
</owl:Class >
```

Figura 39 – Exemplo de utilização da restrição de cardinalidade mínima e máxima
Fonte: Antoniou (2004, p.121)

Buscou-se nessa seção apresentar os elementos principais da OWL para se conseguir construir ontologias formais, isto é, com seus componentes formalmente explícitos: declarações iniciais (*headers*), classes, propriedades e restrições. Informações relacionadas a serviços de inferência em Owl podem ser encontradas em Bechhofer, Horrocks e Patel-Schneider (2003). Informações adicionais sobre a linguagem podem ser encontradas no sitio do W3C[81], cujo complemento enriquece o aprendizado sobre o assunto: propriedades especiais (owl:TransitiveProperty, owl:SymmetricProperty, owl:FunctionalProperty e Owl:InverseFunctional), combinações lógicas para concepção de classes (owl:complementOf, owl:unionOf e owl:intersectionOf), definição de classes através de listas de elementos (owl:oneOf), além da especificidade de cada sublinguagem (OWL Full, OWL DL e OWL Lite).

As Figuras 40 e 41 resumem respectivamente os construtores de classe e os axiomas da OWL.

[81] http://www.w3.org/TR/owl-ref/

Constructor	DL Syntax	Example	Modal Syntax
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male	$C_1 \wedge \ldots \wedge C_n$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer	$C_1 \vee \ldots \vee C_n$
complementOf	$\neg C$	$\neg$Male	$\neg C$
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	{john} $\sqcup$ {mary}	$x_1 \vee \ldots \vee x_n$
allValuesFrom	$\forall P.C$	$\forall$hasChild.Doctor	$[P]C$
someValuesFrom	$\exists P.C$	$\exists$hasChild.Lawyer	$\langle P \rangle C$
maxCardinality	$\leqslant nP$	$\leqslant$1hasChild	$[P]_{n+1}$
minCardinality	$\geqslant nP$	$\geqslant$2hasChild	$\langle P \rangle_n$

Figura 40 – Construtores de classe OWL
Fonte: Bechhofer, Horrocks e Patel-Schneider (2003)

Axiom	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human $\sqsubseteq$ Animal $\sqcap$ Biped
equivalentClass	$C_1 \equiv C_2$	Man $\equiv$ Human $\sqcap$ Male
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male $\sqsubseteq$ $\neg$Female
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{President_Bush} $\equiv$ {G_W_Bush}
differentFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	{john} $\sqsubseteq$ $\neg${peter}
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter $\sqsubseteq$ hasChild
equivalentProperty	$P_1 \equiv P_2$	cost $\equiv$ price
inverseOf	$P_1 \equiv P_2^-$	hasChild $\equiv$ hasParent$^-$
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+$ $\sqsubseteq$ ancestor
functionalProperty	$\top \sqsubseteq \leqslant 1P$	$\top \sqsubseteq \leqslant$1hasMother
inverseFunctionalProperty	$\top \sqsubseteq \leqslant 1P^-$	$\top \sqsubseteq \leqslant$1hasSSN$^-$

Figura 41 – Axiomas OWL
Fonte: Bechhofer, Horrocks e Patel-Schneider (2003)

A linguagem selecionada para o desenvolvimento da ontologia pertence a uma fase específica de seu ciclo de vida. Entretanto, outras fases são previamente necessárias à fase de implementação, de forma que a ontologia possa atender ao seu propósito. Tais fases serão apresentadas na próxima seção.

2.3.3.5 Processos envolvidos na concepção de Ontologias

O objetivo desta seção é apresentar os principais conceitos que cercam o domínio da pesquisa, a fim de facilitar o entendimento do leitor acerca do capítulo sobre o referencial empírico deste trabalho: análise de metodologias para construção de ontologias. Além disso, pretende-se elucidar a relação e a contribuição de áreas como a Engenharia de Software e a Engenharia de Conhecimento, que colaboram, através de seus processos, técnicas e métodos, para um novo ramo de trabalho conhecido como Engenharia ontológica. Considera-se

importante essa abordagem pela possibilidade de composição de técnicas e métodos advindos de tais áreas na proposição de uma metodologia para construção de ontologias, propósito da pesquisa.

Na área de Engenharia de Software existem metodologias[82] já consolidadas, difundidas e bem aceitas na indústria de *software* (PRESSMAN, 2002) como é o caso do *Extreme Programming*, conhecido como XP[83]; e do *Unified Process* (BOOCH; JACOBSON e RUMBAUGH, 2006), ou Processo Unificado, o qual serve de base para a criação de outros processos de *software* devido a sua extensibilidade. Exemplos de processos de *software* baseados no Processo Unificado são o RUP[84] (*Rational Unified Process)*, a sua extensão EUP[85] (*Enterprise Unified Process)* e o Praxis (PAULA FILHO, 2003), ou PRocesso para Aplicativos eXtesíveis e Interativos. O mesmo pode ser dito para a área de Engenharia de Conhecimento, na qual possui metodologias reconhecidas como é o caso da metodologia KADS (WIELINGA, SCHREIBER e BREWKER, 1992), que deu origem a algumas ferramentas de aquisição semi-automática de conhecimento (REZENDE, 2003), e sua extensão, a metodologia CommonKADS[86], utilizada em muitas organizações e universidades Européias para desenvolvimento de sistemas baseados em conhecimento.

Entretanto, percebe-se um cenário relativamente diferente na Engenharia ontológica, na qual diversas metodologias têm sido apresentadas e discutidas para construção, reutilização e avaliação de ontologias (GUARINO e WELTY, 2000; FERNÁNDEZ et al., 1999; JONES; BENCH-CAPON; VISSER, 1998; USCHOLD e GRUNINGER, 1996; GÓMEZ-PEREZ, FERNANDEZ e VICENTE, 1996; GRUNINGER e FOX, 1995), mas apresentam abordagens e características diversas, sendo direcionadas a diferentes propósitos e aplicações, ou seja, não possuem propostas unificadas, sendo que grupos diferentes utilizam diferentes abordagens (FERNÁNDEZ et al., 1999).

Fernández (1999) considera que o processo de desenvolvimento de ontologias deveria ser fundamentado no padrão internacional IEEE-1074, norma para desenvolvimento de *software* advinda da área de Engenharia de Software. O autor ressalta que algumas extensões na norma IEEE-1074 em conjunto com algumas adaptações particulares às ontologias deveriam ser levadas em consideração no processo de construção das mesmas. E justifica a

---

[82] Não se pretende, aqui, realizar uma revisão da literatura sobre os processos de *software* citados.
[83] http://www.extremeprogramming.org/
[84] http://www.ambysoft.com/unifiedprocess/rupIntroduction.html
[85] http://www.enterpriseunifiedprocess.com/
[86] Informações adicionais em http://www.commonkads.uva.nl/

aplicação da norma no processo de construção de ontologias por considerá-las como um componente de produtos de *software*[87].

As seções 2.3.3.5.1 e 2.3.3.5.2 a seguir apresentam respectivamente a norma IEEE-1074 (1997) e atividades envolvidas no processo de construção de ontologias, além de mostrar métodos e técnicas complementares que podem vir a ser úteis em fases específicas do processo de desenvolvimento de sistemas. Vislumbra-se, nessa revisão, elucidar fundamentos que clareiem a questão interdisciplinar que circunda a Engenharia ontológica.

2.3.3.5.1 Processo de Desenvolvimento de *Software* (IEEE-1074)

A norma IEEE-1074 (1997) recomenda padrões na criação de processos de ciclo de vida de *software*, sendo útil a qualquer organização que seja responsável pela gestão e desenvolvimento de projetos envolvendo *software*. A metodologia explicitada na norma seleciona um modelo apropriado para construção de *software* e recomenda uma série de grupos de atividades organizados nas seguintes fases: i) gerência de projeto; ii) pré-desenvolvimento; iii) desenvolvimento; iv) pós-desenvolvimento; e v) processos integrais. Tais fases, juntamente com suas atividades, são descritas a seguir.

Na fase de **gerência de projeto** ocorrem atividades relacionadas ao início de um projeto, ao planejamento da gestão de um projeto e ao monitoramento e controle do projeto de *software* em todo o seu ciclo de vida.

As tarefas de *início de projeto* são focadas na criação e atualização da infra-estrutura do projeto de desenvolvimento ou gerenciamento de um *software*, tais como criação do processo de ciclo de vida do *software*, estimativas de desempenho, alocação de recursos (orçamento, pessoas, equipamentos, recursos computacionais) e definição de métricas (aplicadas aos produtos e processos que afetam o projeto).

As *atividades de planejamento* devem ser direcionadas para toda a gestão do projeto, incluindo planos de contingências. Tais atividades podem possibilitar iterações quando da revisão de uma fase do projeto, e são organizadas da seguinte forma: a) avaliação (inclui auditoria, revisões e testes); b) gestão de configuração (recomenda-se um guia para descrever *como* e *quando* um determinado procedimento é executado); c) transição de sistema (somente nos casos em que sistemas antigos estão sendo substituídos); d) instalação (*hardware* requerido, registro do *software*, tipo de *software* a ser instalado e guia de instalação); e)

[87] "[…] ontologies are part (sometimes only potentially) of *software* products. […] ontologies should be developed according to the standards proposed for *software* generally, which should be adapted to the special characteristics of ontologies."

documentação (produção de documentos para registro do conhecimento acerca do projeto); f) treinamento dos usuários; e g) integração com outros sistemas (caso for necessário).

As *atividades de monitoramento* e *controle* objetivam rastrear e gerenciar todo o projeto, englobando gestão de riscos inerentes ao mesmo (fatores que podem afetar ou prejudicar o alcance dos objetivos), identificação de melhorias para os processos de *software* e organização e tratamento dos registros sobre o projeto (histórico do projeto na organização, como auditorias e planejamento de projetos futuros).

A fase **pré-desenvolvimento** consiste em analisar[88] idéias ou conceitos de um sistema e, em função de problemas observados no ambiente, alocar requisitos para o sistema antes do início de desenvolvimento do *software*. A fase inclui atividades de estudo de viabilidade e análise de requisitos[89] do sistema. O estudo de viabilidade objetiva analisar o custo-benefício para se decidir pela modificação de aplicações existentes ou por um novo desenvolvimento a partir de uma necessidade de solução para algum problema formulado (além de decidir pelo desenvolvimento interno ou pela compra da solução). A justificativa para cada recomendação deve ser documentada e formalmente aprovada por todos os participantes do projeto (incluindo usuários e desenvolvedores). Tal documento pode ser a base de um contrato de desenvolvimento. A análise de requisitos do sistema objetiva fazer um estudo do ambiente atual e produzir um plano de desenvolvimento de *software*, indicando em detalhes os recursos necessários (humanos, materiais, de *software* e de *hardware*), bem como as estimativas de prazos e custos (cronograma e orçamento). Tais requisitos também precisam ser documentados.

A fase de **desenvolvimento** consiste na execução de atividades durante o desenvolvimento de um produto de *software*, a saber: requisitos, projeto e implementação. Tais atividades são conduzidas de acordo com os requisitos funcionais de *software*, *hardware* e pessoas, especificados na etapa pré-desenvolvimento.

As *atividades de requisitos* de *software* abrangem as restrições ou regras que o *software* deve cumprir em função da definição das necessidades do requisitante. Os requisitos devem servir como documento inicial para a realização das tarefas de modelagem e prototipação, e é geralmente iterativa. O uso de uma metodologia abrangente é recomendado para assegurar que os requisitos permaneçam completos e consistentes. Além disso, técnicas advindas da análise estruturada podem auxiliar na atividade de definição de requisitos. Além

---

[88] A análise de alto nível inclui um conjunto de necessidades a nível estratégico e a nível da área de negócios (PRESSMAN, 2002).

[89] Os requisitos dessa etapa não são os mesmos discutidos na etapa de desenvolvimento. São requisitos voltados ao ambiente ou sistema no qual o *software* será instalado.

dos requisitos de *software*, os requisitos de interface do sistema também são demandados nessa etapa, os quais serão determinantes para a usabilidade do sistema (grau de satisfação na interação entre usuário e sistema).

As *atividades de projeto* objetivam desenvolver uma representação bem organizada e coerente do sistema, e que satisfaça os requisitos de *software* especificados nas atividades de requisitos. Tal representação inclui o projeto da arquitetura do *software*, no qual se recomenda o uso de técnicas de modelagem de sistemas (modelagem orientada a objetos, modelagem de dados no caso da aplicação de banco de dados, dentre outras), além da especificação de algoritmos envolvidos e interfaces relacionadas ao usuário, ao *software* e ao *hardware*.

As *atividades de implementação* resultam na transformação da representação do projeto da arquitetura do *software* em uma linguagem de programação. Os códigos fonte e executável resultantes da linguagem empregada serão utilizados para execução de testes de qualidade no *software*, a fim de verificar e validar o produto de acordo com os requisitos demandados.

A fase **pós-desenvolvimento** consiste em atividades de instalação (avaliação da aceitação por parte dos usuários, possíveis modificações), operação e suporte (interação entre o usuário e o sistema, assistência técnica e consultoria quando necessário), manutenção (identificação de problemas, identificação de melhorias podendo ocasionar novas versões) e, quando pertinente, desativação do produto de *software* (substituição do sistema ou atualização de versão).

E, finalmente, a fase de **processos integrais**, utilizada para assegurar que todas as atividades envolvidas no projeto possam ser concluídas com êxito, e, ainda, suas atividades são executadas ao mesmo tempo com atividades dos processos orientados ao desenvolvimento do *software*. Algumas dessas atividades seriam: a) *avaliação* dos processos em todo o ciclo de vida do produto de *software* (condução de revisões e auditorias nos processos, desenvolvimento de procedimento de testes, execução de testes, avaliação de resultados); b) *gerenciamento de configuração* do *software* (controle da configuração de produtos específicos como código do programa, documentação, planos e especificações); c) desenvolvimento e distribuição de *documentação* para desenvolvedores e usuários envolvidos nos processos, a fim de fornecer, em tempo hábil, informações sobre o *software*; e d) atividade de *treinamento* (desenvolver material, validar o programa de treinamento e implementar o programa de treinamento), cujo sucesso do *software* está ligado ao bom conhecimento que as pessoas têm sobre o mesmo.

2.3.3.5.2 Processo de Construção de Ontologias

O ciclo de vida da ontologia contempla um conjunto de estágios e atividades que são utilizados em seu processo de desenvolvimento. O processo de construção de ontologias mais comumente encontrado na literatura (FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997; USCHOLD e GRUNINGER, 1996; NOY e GUINNESS, 2001) seria: i) planejamento; ii) especificação de requisitos; iii) aquisição de conhecimento; iv) conceitualização; v) formalização; vi) integração; vii) implementação; viii) avaliação; e ix) documentação.

Na atividade de **planejamento** ocorre um levantamento das principais tarefas necessárias ao processo de construção da ontologia em função de um estudo do ambiente, descrevendo como tais tarefas serão organizadas, quanto tempo será gasto para cada tarefa, e quais recursos serão demandados no processo, como pessoas, *software* e *hardware*. Além disso, na atividade de planejamento ocorre um estudo de viabilidade buscando responder questões sobre a possibilidade e a adequação de construção da ontologia.

A fase de **especificação de requisitos** requer o conhecimento sobre o domínio que se deseja representar, de forma a identificar possíveis problemas que a ontologia deverá ser capaz de resolver. A definição do propósito da ontologia é demandada nessa fase, a qual determina a necessidade de construção da ontologia e os fins de utilização da mesma. Destaca-se aqui a importância da abstração por parte da pessoa que está executando a atividade de especificação, tendo em vista a subjetividade influenciada por fatores cognitivos e pelo conhecimento prévio de cada indivíduo sobre o domínio. O processo de abstração é responsável pela imposição de coerência acerca dos conceitos capturados como partes essenciais do domínio no qual se pretende representar.

A **aquisição de conhecimento** é uma atividade conhecida na área de Engenharia do Conhecimento, utilizada principalmente na criação de bases de conhecimento para sistemas baseados em conhecimento e sistemas especialistas (REZENDE, 2003). O profissional especializado na execução dessa atividade é o Engenheiro de Conhecimento, que tem como tarefa estudar o domínio e, por meio de uma interação intensa com especialistas da área, criar um modelo que represente as características conceituais do domínio. Na intenção de tornar mais efetivo o processo de aquisição de conhecimento, várias técnicas têm sido desenvolvidas, desde a manual (diretrizes do processo conduzido pelo próprio engenheiro) até a automática (como mineração de conhecimento de extensas fontes de dados ou implantando mecanismos de inferências que permitam aprendizado automático de máquina). As técnicas

manuais, semi-automáticas e automáticas podem ser encontradas em detalhe em Rezende (2003).

Vários métodos de se conseguir uma comunicação precisa entre o especialista do domínio e o profissional envolvido na análise e desenvolvimento de sistemas têm sido propostos ao longo dos anos pela Engenharia de Software e Análise de Sistemas. Booch, Jacobson e Rumbaugh (2006) inserem os casos de uso como formas de capturar o conhecimento obtido na interação com os usuários e mostrar (através de texto livre ou diagrama) como os atores (usuário, outro sistema) interagem com as funções principais do sistema de informação. Segundo os autores, os casos de uso permitem verificar se o desenvolvedor e o usuário concordam sobre o que o sistema deve fazer. Costa (1994)[90] apud Almeida (2006) inseri o método JAD – *Joint Application Design* – como um recurso de discussões em grupo ao invés de entrevistas individuais. As decisões sobre o projeto são praticadas em consenso entre os participantes, tornando o processo interativo, viabilizando, dessa forma, a comunicação e o entendimento dos fatos entre as pessoas envolvidas no processo. Um outro método é a *análise de cenários* (BENNER et al., 1993), que possibilita uma comunicação efetiva entre as pessoas envolvidas no processo através da simulação da realidade de trabalho de cada usuário, ou seja, descreve o uso de um tipo de sistema para cada usuário envolvido.

No campo das ontologias, o profissional responsável pelo processo de construção, conhecido como *ontologista*, também necessita estudar o ambiente e criar um modelo conceitual do domínio. Dessa forma, a atividade de aquisição de conhecimento é também praticada pela Engenharia ontológica na obtenção de conhecimento sobre o domínio. O processo de aquisição de conhecimento ocorre, muitas das vezes, simultâneo a outras atividades, como nas atividades de especificação e conceitualização (FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997). As possíveis fontes para obtenção de conhecimento são: especialistas do domínio, livros, manuais, figuras, tabelas, outras ontologias, etc. E as possíveis técnicas para se obter conhecimento são: entrevistas não-estruturadas (conhecida como técnica de *brainstorming*) e estruturadas, ferramentas de aquisição de conhecimento e análise informal e formal de documentos. Uma técnica que auxilia na análise e apreensão do conteúdo dos documentos é a *análise de assunto*, originada na biblioteconomia. Segundo Fujita (2003), a análise de assunto permite identificar e selecionar os conceitos que representam a essência do documento. A identificação de conceitos, realizada durante a

---

[90] COSTA, W.D. JAD-Joint Applicatin Design: como projetar sistemas de informação mais eficientes. Rio de Janeiro: IBPI, 1994. 115p.

leitura do documento, é considerada uma atividade intelectual importante no processo de indexação, que irá implicar na correspondência precisa com o assunto pesquisado em índices.

A fase de **conceitualização** descreve o problema e sua solução através de um modelo conceitual que representa o domínio. Constrói-se uma lista de termos, incluindo conceitos, verbos, instâncias e propriedades, que busca identificar e reunir o conhecimento útil do domínio. Uma dificuldade encontrada nessa fase é na definição da terminologia direcionada aos conceitos e as relações. Pode acontecer de um mesmo termo ser utilizado para representar vários conceitos, o que pode gerar ambigüidades na interpretação do conhecimento. Outras vezes, vários termos representam um mesmo conceito. Tais problemas podem propiciar modelos incompletos e inconsistentes, tendo em vista que o modelo ainda encontra-se no nível informal, isto é, suas definições são apresentadas em linguagem natural, dando margem à existência de ambigüidades e inconsistências.

É na fase de conceitualização que os conceitos abstraídos na fase de especificação são organizados em uma taxonomia resultando em classes conceituais. Na composição das classes, a atividade de categorização é fundamental, pois fornece ao desenvolvedor meios de analisar os conceitos do domínio e agrupá-los quando da observância de semelhança entre eles. Existem diferentes técnicas de raciocinar sobre o domínio a fim de se construir a taxonomia da ontologia. A técnica de *bottom-up* refere-se ao processo de construção de um modelo no qual o conhecimento já elicitado é generalizado; a técnica de *top-down* consiste em construir um modelo abstrato e posteriormente especializá-lo a partir do conhecimento do domínio. E a técnica de *middle-out* que identifica os conceitos nucleares do domínio e posteriormente especializa e generaliza tais conceitos de maneira simultânea. Finalmente, após a construção da estrutura da ontologia, os conceitos que se enquadram (de acordo com as restrições de propriedades) nas classes definidas são classificados e denominados instâncias ou objetos de tais classes conceituais.

A fase de **formalização** consiste em transformar o modelo conceitual concebido na fase de conceitualização em um modelo formal a fim de definir de forma precisa o seu significado. O profissional envolvido na construção da ontologia concentra-se no processo de modelagem computacional do problema, usando, por exemplo, a lógica de primeira ordem e suas extensões (sistemas de representação baseados em *frame*, lógica descritiva, etc). As técnicas empregadas nessa fase são oriundas da área de Inteligência Artificial, conforme foram discutidas na seção 2.3.1. A IA desenvolve linguagens capazes de expressar o conhecimento de modo claro, evitando múltiplas interpretações para um mesmo fato. A formalização também permite ao desenvolvedor tratar questões de complexidade

computacional do problema, isto é, alguns problemas resolvidos através de uma linguagem formal são computacionalmente intratáveis (conforme foi citado na seção sobre owl e ontologias, na qual se discutiu sobre os problemas da owl-full). Essa complexidade de processamento fica explícita na formalização dos conceitos relevantes para o problema (REZENDE, 2003). Daí a importância do profissional saber escolher a linguagem de representação que melhor se enquadre nas necessidades da ontologia.

Uma vez feita a formalização, a ontologia deve ser codificada em uma linguagem formal de forma a ser legível para a máquina. A fase de **implementação** consiste em mapear o modelo formal em uma linguagem que se adeque às demandas como Prolog[91], Ontolingua[92], RDFS, OWL, etc. Uma alternativa para facilitar a construção de ontologias é buscar **integrar** a ontologia em questão a ontologias existentes. Esta fase considera a reutilização de conceitos existentes em outras ontologias. A proposta é examinar a conceitualização de meta-ontologias (chamadas ontologias de alto nível) e selecionar (parcialmente ou por completo) àquelas que melhor se ajustarem ao modelo que está sendo construído. Ferramentas para construção de ontologias podem ser utilizadas como apoio, pois a fase de implementação é considerada dispendiosa. Uma lista de ferramentas e linguagens para construção de ontologias, juntamente com a descrição das mesmas, pode ser encontrada em Almeida e Bax (2003, pgs. 15-16).

Após a construção da ontologia é necessário verificar se ela atende ao propósito para o qual foi desenvolvida. A fase de **avaliação** consiste em executar um julgamento técnico da ontologia, isto é, verificar se o conhecimento apreendido na estrutura corresponde ao domínio no qual foi feita a aquisição de conhecimento. Almeida (2006) apresenta algumas abordagens para validação de conteúdo da ontologia advindas da Ciência da Informação, como qualidade da informação para estudar usabilidade, necessidades de usuários e de uso da informação; qualidade de dados, proveniente da Ciência da Computação; questões de competência, da área de pesquisa de ontologias; e objetivos educacionais de aprendizado, da área de Educação. Algumas propostas de metodologias para avaliação de ontologias podem ser encontradas em Guarino e Welty (2000) e Gómez-Pérez (1994).

E, finalmente, a atividade de **documentação**, considerada imprescindível em cada fase do ciclo de vida da ontologia no sentido de registrar todo o conhecimento acerca do projeto de construção. Pode vir a ser útil no caso de surgir a necessidade de **manutenção** (modificação na estrutura) e reuso da ontologia em outras aplicações.

---

[91] É uma linguagem de programação que se enquadra no paradigma de programação em lógica matemática. Geralmente está associada à área de inteligência artificial.
[92] Linguagem que combina paradigmas de linguagens baseadas em *frames* e lógica de primeira ordem.

Após as apresentações de um possível processo de construção de ontologias e do processo de *software* advindo da norma IEEE-1074 (1997), constata-se a semelhança entre algumas fases e atividades advindas de ambos os casos. Atividades como planejamento, modelagem, implementação, avaliação e documentação são recomendadas nas duas abordagens, tendo em vista algumas particularidades inerentes às ontologias. Algumas técnicas empregadas na atividade de aquisição de conhecimento da Engenharia de Conhecimento são também contempladas nas atividades de especificação de requisitos e modelagem da Engenharia de Software, bem como da Engenharia ontológica. Segundo Alonso et al. (1996)[93] apud Fernandez, Gomez-Perez e Juristo (1997, p.35) "[...] modelos de ciclo de vida são descritos na Engenharia de Software para construção de sistemas e transferidos para a Engenharia de Conhecimento" [94]. Senso assim, o ciclo de vida da ontologia pode ser associado ao ciclo de vida de um *software* como acontece em algumas propostas metodológicas para construção de ontologias: a Methontology (FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997) e a metodologia baseada no projeto Tove – *Toronto Virtual Enterprise Ontology* (GRUNINGER e FOX, 1995), conforme poderá ser visto no capítulo 4 – Apresentação da análise dos objetos empíricos.

Finalmente, as áreas elucidadas nessa seção fundamentam-se basicamente em *processos* (atividades, métodos, técnicas, práticas e transformações), *tecnologia* (linguagens de desenvolvimento, ferramentas de apoio ao desenvolvimento, bancos de dados, bases de conhecimento) e *pessoas* (clientes, usuários finais e equipe de projeto). Sendo assim, as ontologias, vistas como produto de *software*, podem ser construídas a partir de métodos e técnicas advindos da Engenharia de Software e Engenharia do Conhecimento.

No capítulo 4, sobre o referencial empírico desta pesquisa, são apresentados maiores detalhes sobre o processo de construção de ontologias, baseando-se em trabalhos de outros autores da área de ontologias, e em métodos e técnicas complementares advindas das áreas de Engenharia de Software e Engenharia de Conhecimento.

---

[93] ALONSO, F.; JURISTO, N.; MATÉ, J.L; PAZOS, J. Software Engineering and Knowledge Engineering: Towards a Common Life-Cycle**.** *Journal of Systems and Software*. N.33.1996. Pags 65-79.

[94] "[...] life cycle models have been described in Software Engineering and transferred to Knowledge Engineering."

3 - Metodologia de Pesquisa

O objetivo geral deste trabalho é propor princípios metodológicos para a construção de ontologias de domínio para representação do conhecimento em sistemas de informações, através da análise da literatura sobre metodologias para construção de ontologias, das normas internacionalmente aceitas para construção de vocabulários controlados, observando-se também as normas amplamente adotadas para construção de *software*. Após a conclusão da fase de fundamentação teórica da presente pesquisa, iniciou-se em novembro de 2007 as etapas de: **i**) identificação e seleção de documentos que contemplassem o assunto sobre metodologias para construção de ontologias; **ii**) identificação e seleção de metodologias para construção de ontologias em tais documentos; **iii**) seleção de metodologias para construção de vocabulários controlados; **iv**) definição das categorias de análise das metodologias, dos métodos e da norma, de modo a coletar os dados pertinentes à pesquisa; **v**) análise comparativa das metodologias, dos métodos e da norma; e **vi**) desenvolvimento da proposta metodológica para construção de ontologias.

A fundamentação teórica foi essencial para a formação de um arcabouço teórico sobre ontologias, bem como de áreas interdisciplinares relacionadas à temática representação do conhecimento, o que contribuiu para o alcance do objetivo geral da pesquisa.

Para a identificação de documentos, foram consultadas bases de dados de documentos científicos e empregada a técnica de busca por palavras-chave que refletissem o universo do assunto. O processo de análise e interpretação dos documentos recuperados se deu através da utilização de técnicas de análise de conteúdo, as quais permitiram a seleção dos documentos. Após a obtenção da amostragem documental, pôde-se, finalmente, identificar as metodologias discutidas nos documentos e justificar quando da inserção de uma delas no material empírico destinado à análise. Tal identificação também foi realizada através de técnicas de análise de conteúdo.

A análise de conteúdo contempla um conjunto de técnicas de análise das comunicações visando obter indicadores (quantitativos ou não) que permitam a inferência de conhecimentos presentes nas mensagens (BARDIN, 1977). O emprego de tais técnicas é descrito nas seções 3.2 e 3.3.

As etapas de análise do material empírico e de análise comparativa do mesmo, iniciadas em fevereiro de 2008, indicaram a viabilidade da pesquisa, isto é, apontaram a viabilidade de uma proposta metodológica para construção de ontologias. E, finalmente, a

etapa de desenvolvimento da proposta metodológica para construção de ontologias, iniciada em abril de 2008, foi conduzida dando ênfase a padrões metodológicos e a um conjunto de teorias advindas de campos-chave desta pesquisa, a **Ciência da Informação** e a **Ciência da Computação**, tendo as áreas de Engenharia do Conhecimento e Engenharia de Software representado papéis fundamentais neste último campo. Diante disso, a proposta buscou responder à questão de pesquisa "*Será possível confirmar a possibilidade de proposição de princípios metodológicos para construção de ontologias, à luz das contribuições pertinentes das Ciências da Informação e da Computação, dentre outros campos do conhecimento?*".

O presente capítulo apresenta a descrição dos passos metodológicos que tornaram possível a concretização deste trabalho de pesquisa. Na seção 3.1 são apresentadas as bases de dados, que determinaram a recuperação dos documentos pesquisados. Na seção 3.2 são descritos os passos e os critérios utilizados na seleção das metodologias para construção de ontologias e das metodologias para construção de vocabulários controlados, ambos compondo o material empírico para análise. Na seção 3.3 são justificadas as categorias de análise do material empírico. Além disso, são descritos ainda nesta seção os instrumentos que foram utilizados na estruturação da coleta de dados e no registro dos dados. A seção 3.4 apresenta como foi conduzida a atividade de análise comparativa dos materiais envolvidos na pesquisa. E, finalmente, a seção 3.5 apresenta a forma como foi conduzido o trabalho de proposição da metodologia para construção de ontologias, a partir do esboço comparativo descrito na seção 3.4.

3.1 Seleção das fontes de pesquisa

Considerando sua amplitude, não seria possível investigar, no escopo de uma pesquisa de mestrado, toda a bibliografia relacionada com o tema de pesquisa, isto é, *metodologias para construção de ontologias*, assim como realizar uma investigação exaustiva em fontes de publicações científicas. Portanto, decidiu-se utilizar, como material empírico, artigos científicos e livros que tivessem como objetivo apresentar pesquisas sobre o tema em pauta.

As fontes de obtenção dos artigos científicos foram o portal de periódicos da Capes[95] e a biblioteca digital Citeseer[96], sendo que ambas as fontes disseminam literatura científica nos campos da Ciência da Informação e da Ciência da Computação. Além disso, proporcionam análises estatísticas das citações para todos os documentos do repositório, o que possibilita ao

[95] http://www.periodicos.capes.gov.br/portugues/index.jsp
[96] http://citeseer.ist.psu.edu

pesquisador tomar conhecimento da opinião dos pares em relação ao artigo em questão e, em conseqüência, de sua aceitação pela comunidade. A indexação das citações no Citeseer é feita de forma autônoma, ou seja, sem intervenção humana (LAWRENCE, BOLLACKER e LEE GILES, 1999). Quando da execução deste estudo, foi constatada a atualização da base de dados do Citeseer até o período de setembro de 2006[97].

Os livros sobre construção de ontologias foram pesquisados no catálogo de bibliotecas da Universidade Federal de Minas Gerais (UFMG) através do sistema Pergamum[98], seguindo o critério de maior disponibilidade.

3.2 Seleção do material empírico

A presente seção apresenta os passos que determinaram a seleção do material empírico nesse estudo e se organiza da seguinte forma: na seção 3.2.1 é descrita a estratégia de pesquisa utilizada na atividade de seleção das metodologias para construção de ontologias; e a seção 3.2.2 apresenta os critérios adotados na seleção das metodologias para construção de vocabulários controlados.

3.2.1 Seleção das metodologias para construção de ontologias

Na atividade de seleção das metodologias para construção de ontologias, foi realizada uma investigação na literatura das áreas da computação, como também da Ciência da Informação, tendo-se mais sucesso na primeira. Tal fato pode ser explicado devido às ontologias de sistemas de informações serem objetos de estudo proeminentes da Ciência da Computação.

A estratégia de pesquisa para a seleção das metodologias se deu pelos seguintes passos, que serão detalhados nas próximas subseções:

1. Seleção de artigos e livros que tratam do assunto “metodologias para construção de ontologias” através da recuperação por palavras-chave determinadas;
2. Identificação de metodologias para construção de ontologias nos documentos selecionados;

---

[97] http://citeseer.ist.psu.edu/source.html
[98] http://webpergamum.adm-serv.ufmg.br/biblioteca/php/opcoes.php

3. Quantificação de ocorrências de citação das metodologias identificadas nos documentos;
4. E, finalmente, determinação das metodologias que foram analisadas comparativamente na pesquisa.

É válido destacar neste ponto a diferença entre os termos *metodologia* e *método*, visto que na investigação realizada na literatura, ambos os termos são usados por vezes de maneira indiscriminada. Contudo, constatou-se que alguns dos objetos a serem identificados e analisados são considerados metodologias, outros são considerados métodos, conforme será visto ainda neste capítulo e mais adiante no capítulo 4 – apresentação da análise dos objetos empíricos.

Segundo definições do IEEE (1990)[99] apud Fernandez, Gómez-Perez e Corcho (2004), uma *metodologia* seria "uma série integrada de técnicas ou métodos criando uma teoria geral de sistemas de como uma classe de pensamento pode ser executada" [100]. Um *método* "é um conjunto de processos ou procedimentos ordenados usados na engenharia de um produto ou na realização de um serviço"[101], tendo-se em vista que tais processos são compostos de atividades, que, por sua vez, são compostas por tarefas atribuídas a um ou mais membro do projeto. Geralmente, tarefas relacionadas são agrupadas para formar atividades. E uma técnica seria "um procedimento técnico e gerencial usado para alcançar um dado objetivo"[102], que em outras palavras significa dizer um modo pelo qual o método é executado. A Figura 42 ilustra graficamente as relações entre métodos, processos, atividades, tarefas e técnicas em uma metodologia.

---

[99] IEEE *Standard Glossary of Software Engineering Terminology*. IEEE Computer Society. New York. IEEE Std 610.121990, 1990.

[100] "[…] integrated series of techniques or methods creating a general systems theory of how a class of thought […] be performed."

[101] "[…] is a set of orderly process or procedure used in the engineering of a product or performing a service."

[102] "[…] a technical and managerial procedure used to achieve a given objective"

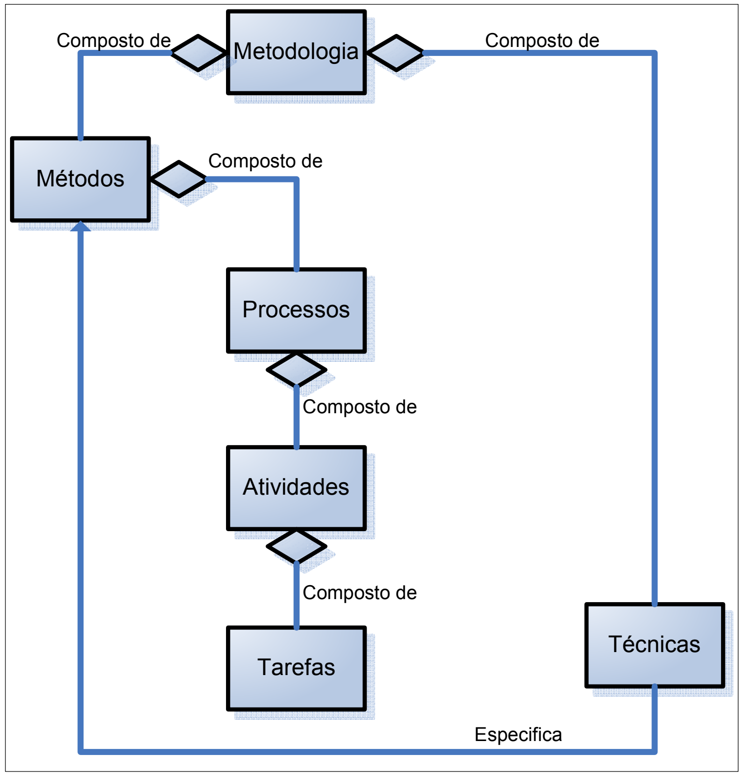


Figura 42 – Representação gráfica de uma metodologia
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004, p.109)

Segundo de Hoog (1998)[103] apud Fernandez, Gómez-Perez e Corcho (2004), metodologia e método são conceitos distintos, pois uma metodologia refere-se a conhecimento sobre métodos, isto é, determina "como" e "quando" uma dada atividade pode ser realizada. Sendo assim, uma metodologia é composta de métodos que possuem suas próprias técnicas.

3.2.1.1 Passo 1 – Seleção dos documentos

O primeiro passo foi selecionar os artigos científicos que fossem passíveis de análise para levantamento das metodologias discutidas na literatura. O processo de recuperação de artigos na biblioteca Citeseer se dá por palavras-chave ou através de *links*[104]de citação (LAWRENCE, BOLLACKER e LEE GILES, 1999). A biblioteca oferece uma opção de busca de seu conteúdo através do mecanismo de busca do Google (denominada *Google Docs*). Essa estratégia de busca é interessante, pois, para uma maior precisão, o Google pesquisa exatamente as palavras ou expressões que forem indicadas no campo de pesquisa

---

[103] de HOOG R. *Methodologies for Building Knowledge Based Systems*: Achievements and Prospects. In: Liebowitz J (ed) Handbook of Expert Systems. Chapter 1, Boca Raton, Florida: CRC Press. 1998.
[104] "CiteSeer allows the location of papers by keyword search or by citation links"

através de aspas. O Google também adiciona automaticamente um "e" entre as palavras inseridas na expressão de pesquisa. O resultado mostra páginas onde ocorram todas as palavras indicadas na busca.

Diante disso, as palavras-chave precisaram ser determinadas de forma a representar o assunto, ou seja, "metodologias para construção de ontologias". As palavras-chave para a atividade de recuperação foram então escolhidas em função do conhecimento prévio adquirido na leitura de documentos sobre o assunto: "*methodology*", "*methodologies",* "*method*", "*design*", "*building*", "*development process*", "*life cycle*", em conjunto com "*ontologies*" ou "*ontology*". É importante salientar que as palavras-chave estão em inglês em virtude dos materiais publicados sobre ontologias se encontrarem, em sua maioria, nesta língua.

A consulta montada com a expressão "*methodology for building ontologies*" resultou na recuperação de 70 páginas no *Google Docs*, já "*methodologies for building ontologies*" resultou na recuperação de 22 páginas. Além da estratégia de busca por expressão, foi pesquisada cada palavra-chave em conjunto com "*ontologies*" ou "*ontology*". Na análise de busca das palavras-chave, constatou-se que a palavra "*ontology*" em conjunto com as outras palavras-chaves foi a que resultou em maior retorno de páginas, conforme pode ser visto nas Tabelas 7 e 8. Na análise realizada nos dois contextos, percebeu-se uma coincidência nas páginas recuperadas, isto é, a mesma página era retornada tanto com "*ontologies*" quanto "*ontology*", constatando uma variação no uso dos termos nos artigos.

Do universo de páginas recuperadas, foram investigadas as 30 primeiras páginas, pelo fato de, na maioria dos casos, as demais páginas retornarem documentos repetidos ou fora do escopo, tendo em vista que os resultados são listados em ordem de importância[105]. Foram selecionados os documentos que mais se aproximavam do objeto de investigação. Para efetuar a escolha, foi necessário analisar o título mencionado na página, fazer uma leitura do resumo (*abstract*) do documento[106] e analisar as palavras-chave do mesmo. Além disso, para uma compreensão macro do objeto de investigação, fez-se uma breve leitura da seção envolvida com o assunto. Tal leitura foi embasada em uma técnica da análise de conteúdo conhecida como *leitura flutuante*, na qual consiste em estabelecer contato com os documentos a serem analisados, obtendo impressões e orientações acerca do conteúdo (BARDIN, 1977, pgs. 96-

[105] Informações sobre o *PageRank* em http://www.google.com/technology/

[106] A página recuperada no *Google Docs* remete ao documento indexado no Citeseer. A biblioteca disponibiliza um recurso de *download* para visualização do documento em vários tipos de arquivo, conforme pode ser visto na Figura 43.

98). Algumas regras advindas da análise de conteúdo também foram consideradas na seleção dos documentos e são descritas a seguir:

- Regra da exaustividade: "*não se pode deixar de fora qualquer elemento por esta ou por aquela razão que não possa ser justificável no plano do rigor*". Muitos dos documentos filtrados não abordavam diretamente o tema "metodologias para construção de ontologias", mas outros aspectos relacionados, tais como "ferramentas", "linguagens", "aplicações", "engenharia", dentre outros, que poderiam servir para a análise.
- Regra da representatividade: "*a amostragem diz-se rigorosa se a amostra for parte representativa do universo inicial*". Neste caso, representando os documentos que tratassem diretamente de metodologias para construção de ontologias.

Além das técnicas expostas acima, foram levados em consideração na análise de cada artigo: a) o número de citações recebidas e b) citação de autores mais produtivos na área de ontologias como Nicolas Guarino, Thomas Gruber, Fernandez Lopes, Asuncion Gómez Perez, Mike Uschold, Michael Gruninger, dentre outros. A Figura 43 mostra um artigo selecionado para análise com o número de citações recebidas. A Figura 44 mostra a listagem dos artigos citados no mesmo, juntamente com um gráfico mostrando a distribuição das citações em função do ano de publicação dos artigos citados. Pôde-se notar, no universo pesquisado, um aumento de citações a partir da década de 90, mostrando que o tema "ontologias" ganhou notoriedade na ciência da computação neste período.

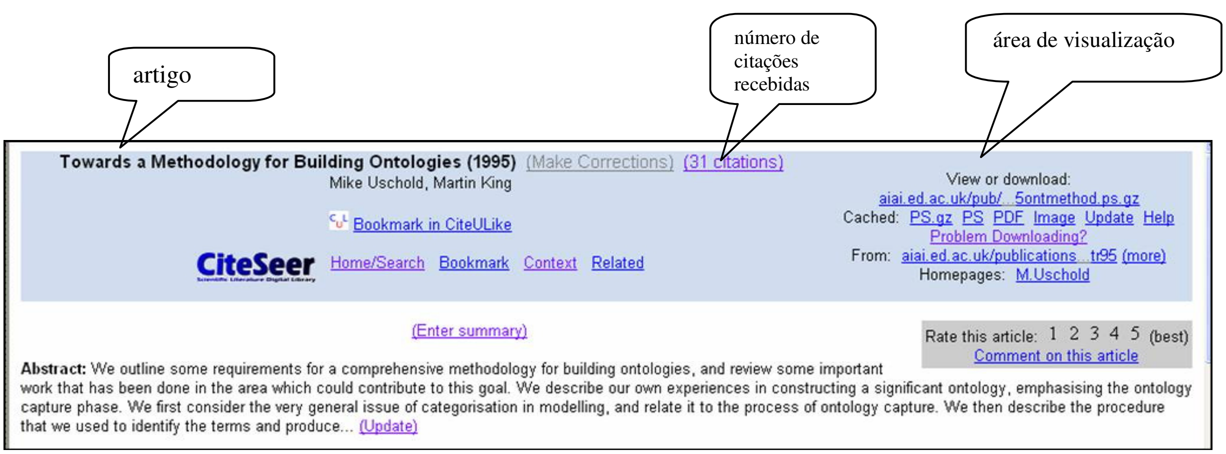


Figura 43 – Número de citações recebidas pelo artigo

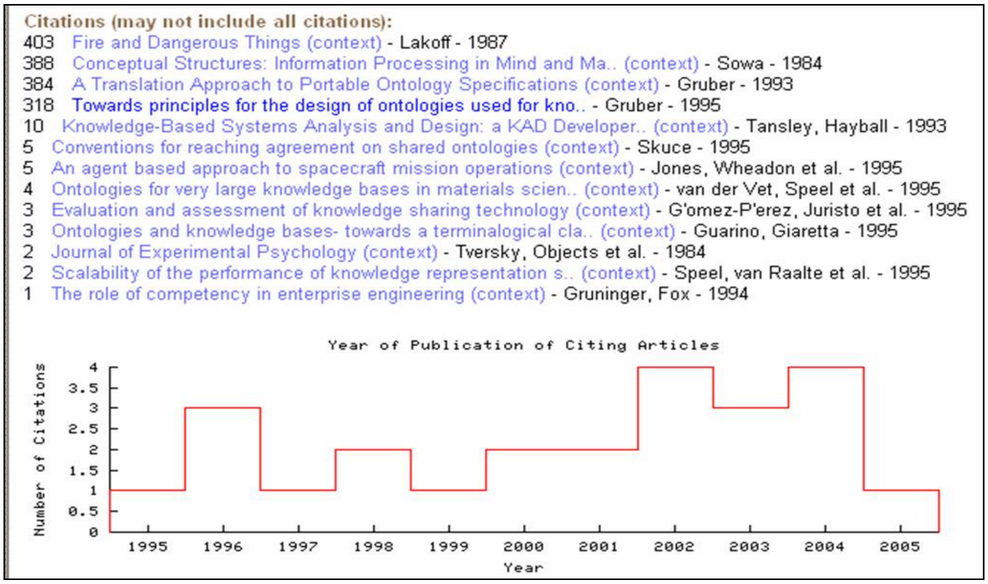


Figura 44 – Número de citações em função do ano de publicação dos artigos

As Tabelas 7 e 8 a seguir mostram o resultado da pesquisa no Citeseer através do *Google Docs*. O resultado de palavras-chave com nenhuma seleção de documentos se deu pelas seguintes razões: a) documentos já haviam sido selecionados por outras palavras-chave; b) assuntos fora do escopo, mostrando que a palavra-chave determinada não era tão relevante na busca, como nos casos das palavras "*design*", "*life cycle*" e "*development process*".

Expressões e palavras-chave	Resultado da busca (páginas)	Docs. selecionados nas primeiras 30 págs.
"methodology for building ontologies"	70	6
"methodologies for building ontologies"	22	2
"methodologies" e "ontologies"	334	1
"methodology" e "ontologies"	1220	4
"method" e "ontologies"	1450	1
"design" e "ontologies"	3140	0
"building" e "ontologies"	2310	0
"development process" e "ontologies"	66	0
"life cycle" e "ontologies"	70	0

Tabela 7 – Resultado da pesquisa com a palavra-chave *ontologies*

Palavras-chave	Resultado da busca (páginas)	Docs. selecionados nas primeiras 30 págs.
"methodologies" e "ontology"	450	1
"methodology" e "ontology"	1780	3
"method" e "ontology"	2560	0
"design" e "ontology"	4980	0
"building" e "ontology"	3580	1
"development process" e "ontology"	95	0
"life cycle" e "ontology"	99	0

Tabela 8 - Resultado da pesquisa com a palavra-chave *ontology*

É válido ressaltar que a partir dos documentos selecionados foi possível identificar novos conceitos e autores para realização de novas buscas no Citeseer, bem como em editoras específicas do portal da Capes, que permitiram a identificação de outros artigos relevantes para a pesquisa. Tais artigos também foram considerados na análise de identificação de metodologias para construção de ontologias e, quando considerados relevantes, foram inseridos na amostra de documentos.

Dando continuidade ao processo de seleção de documentos, o próximo passo foi recuperar livros que tratassem de metodologias para construção de ontologias. Para tal, foi acessado o sistema de gestão de acervos das bibliotecas da UFMG (Pergamum). As referências foram coletadas a partir de bases secundárias que refletem a literatura nacional e internacional nos campos de Ciência da Informação e Ciência da Computação. Investigou-se o acervo *on-line* das bibliotecas da Escola de Ciência da Informação (ECI) e do Instituto de Ciências Exatas (ICEX), por serem considerados campos mais pertinentes em relação ao assunto, e para diminuir o "ruído" causado pelas acepções de ontologia no campo da filosofia. As palavras-chave determinadas para a busca foram: "ontologia", "ontologias", "*ontology*", "*ontologies*", "*engineering*" e "*web*".

Foram recuperadas obras que tratavam de assuntos relacionados a comércio eletrônico, web semântica e recuperação de informação. Foi constatado pouco volume de obras relacionadas à temática "ontologias de sistemas de informações" nos acervos da ECI, bem como nos do ICEX, evidenciando que o assunto de pesquisa ainda é recente nas duas instituições. Foi feita uma análise no sumário de cada obra a fim de verificar a existência do assunto investigado.

Após a realização da atividade de seleção dos documentos descrita acima, foi obtida uma amostra de 25 (vinte e cinco) documentos representativos de análise. Foram selecionados

23 (vinte e três) artigos científicos e 2 (dois) capítulos de livros. A Tabela 9 (a) mostra a relação desses documentos por ordem crescente de ano de publicação. As referências completas desses documentos são apresentadas no Anexo A.

Referência	Autoria	Título	Assunto abordado	Publicação
R1	Mark S. Fox	The TOVE Project: towards a common-sense model of the enterprise	Ontologia aplicada a modelo de negócio	1992
R2	Michael Gruninger e Mark S. Fox	The Design and Evaluation of Ontologies for Enterprise Engineering	Projeto e avaliação de ontologies para engenharia de empresas	1994
R3	Michael Gruninger e Mark S. Fox	Methodology for the design and evaluation of ontologies	Metodologia para construção de ontologia no domínio de negócio	1995
R4	Mike Uschold e Martin King	Towards a Methodology for Building Ontologies	Metodologia para construção de ontologias	1995
R5	J. Bouaud, B. Bachimont, J. Charlet e P. Zweigenbaum	Methodological Principles for Structuring an "Ontology"	Princípios metodológicos para estruturação de uma ontologia	1995
R6	Asuncion Gómez-Perez, Mariano Fernandez-Lopes e Antônio J.de Vicente	Towards a Method to Conceptualize Domain Ontologies	Processo de conceitualização de conhecimento no domínio da química	1996
R7	Michael Gruninger e Mike Uschold	Ontologies: principles, methods and applications	Metodologia para construção de ontologia no domínio de negócio	1996
R8	Mike Uschold	Building ontologies: towards a unified methodology	Metodologia para construção de ontologias para domínio de negócio	1996
R9	Michael Gruninger e Mark S. Fox	The Logic of Enterprise Modelling	Modelagem de empresas	1996
R10	Asuncion Gómez-Perez, Mariano Fernandez-Lopes e Natalia Juristo	Methontology: from ontological art towards ontological engineering	Ciclo de vida de uma ontologia; processo de desenvolvimento; engenharia de software	1997
R11	Michael Gruninger e Mark S. Fox	On Ontologies and Enterprise Modelling	Modelagem de empresas	1997
R12	Hafner, Carole D.; Noy, Natalya Fridman	The state of the art in ontology design: a survey and comparative review	O estado da arte sobre projetos de ontologias	1997
R13	Mike Uschold, Martin King, Stuart Moralee e Yannis Zorgios	The Enterprise Ontology	Ontologia aplicada a modelo de negócio	1998
R14	Dean Jones, Trevor-Capon e Pepijn Visser	Methodologies for ontology development	Metodologias para construção de ontologias	1998
R15	Mariano Fernandez Lopes	Overview of methodologies for building ontologies	Metodologias para construção de ontologias	1999
R16	Dean M. Jones e Ray C. Paton	Toward Principles for the Representation of Hierarchical Knowledge in Formal Ontologies	Representação de relacionamentos hierárquicos na ontologia	1999
R17	Yannis Kalfoglou	Deploying Ontologies in Software Design	Utilização de ontologies em projetos de softwares	2000

Tabela 9 (a) – Relação dos documentos selecionados

Referência	Autoria	Título	Assunto abordado	Publicação
R18	York Sure, Juergen Angele e Steffen Staab	OntoEdit: Guiding Ontology Development by Methodology And Inferencing	Desenvolvimento de ontologia na ferramenta OntoEdit	2002
R19	York Sure, Steffen Staab e Rudi Studer	Methodology for Development and Employment of Ontology based Knowledge Management Applications	Metodologia para desenvolvimento de ontologias em aplicações de gestão do conhecimento	2002
R20	York Sure e Rudi Studer	On-To-Knowledge Methodology Final Version	Metodologia On-To-Knowledge para soluções de gestão do conhecimento em empresas	2002
R21	Robert Stevens, Carole Goble, Ian Horrocks and Sean Bechhofer	Building a Bioinformatics Ontology Using OIL	Desenvolvimento de ontologia de bioinformática usando a linguagem OIL	2002
R22	York Sure, Steffen Staab e Rudi Studer	On-To-Knowledge Methodology (OTKM)	Metodologia On-To-Knowledge para soluções de gestão do conhecimento em empresas	2003
R23	Asuncion Gómez-Perez, Mariano Fernandez-Lopes e Oscar Corcho	Methods and Methodologies for building ontologies. Capítulo do livro: Ontological Engineering	Engenharia de Ontologias; Metodologias para construção de ontologias	2004
R24	Karin Breitman	Por onde começar? Metodologias para a construção de ontologias: Arte ou Engenharia? Capítulo do livro: Web Semântica - A Internet do Futuro	Web Semântica; Metodologias de construção de ontologias	2005
R25	Jorge Cardoso	The Semantic Web Vision: Where are We?	Visão geral da Web Semântica: linguagens, ferramentas, metodologias.	2007

Tabela 9 – Relação dos documentos selecionados

3.2.1.2 Passo 2 – Identificação das metodologias para construção de ontologias

O segundo passo foi identificar as metodologias para construção de ontologias abordadas nos documentos explicitados na Tabela 9 (a).

No passo de seleção dos documentos pôde-se perceber uma carência de metodologias representativas. Segundo a literatura pesquisada, do ponto de vista metodológico, não há um padrão amplamente aceito para construção de ontologias (JONES; BENCH-CAPON e VISSER, 1998; FERNÁNDEZ, GÓMEZ-PEREZ e JURISTO, 1997; USCHOLD e GRUNINGER, 1996). Este fato demonstra que “metodologia para construção” é um assunto

ainda em amadurecimento. Uschold e Gruninger (1996) comentam da dificuldade em encontrar na literatura da área, assuntos sobre “metodologias para construção de ontologias”. Segundo os autores “[...] não há muitas publicações na área, mesmo na literatura pesquisada” [107]. Mesmo em função de essas afirmações terem sido feitas a cerca de dez anos, não foi constatado nessa pesquisa um avanço em relação a propostas de metodologias para construção de ontologias. Isto pode ser conferido no mapeamento temporal[108] das metodologias abordadas nos documentos (1992 até 2007) e em publicações recentes envolvendo o assunto (FERNÁNDEZ, GÓMEZ-PEREZ e CORCHO, 2004; BREITMAN, 2005; CARDOSO, 2007).

Todos os fatos expostos anteriormente, além de frisar a importância desta pesquisa, reforçam o argumento de que a fase de seleção das metodologias para construção de ontologias apresentou dificuldades. Os motivos foram os seguintes: a) as metodologias discutidas na literatura eram quase sempre as mesmas; e b) poucos trabalhos publicados sobre o tema. Diante disso, julgou-se pertinente, em termos de obtenção de metodologias representativas de análise, obter a freqüência de ocorrência de citação nos documentos selecionados. Através da tabela de freqüência de ocorrência tornar-se-ia possível identificar as metodologias de construção mais discutidas e dar ênfase na metodologia que mais se destacou na literatura.

É válido ressaltar que alguns dos documentos forneciam poucas informações sobre o processo de construção de ontologias. Estes discutiam sobre ontologias de alto nível e ontologias de domínio (CHANDRASEKARAN, JOHNSON e BENJAMINS, 1999), porém não explicitavam seus métodos ou metodologias de construção (JONES; BENCH-CAPON e VISSER, 1998). Já outros apontavam metodologias que tratavam de avaliação de taxonomia (GUARINO e WELTY, 2000) e integração (JONES e PATON, 1999) de ontologias. Foram, portanto, mapeadas apenas metodologias e métodos que contemplassem o processo de construção de uma ontologia, conforme pode ser conferido na Tabela 10.

---

[107] “[…] nor is there much published in this área, even in the research literature”.
[108] A Tabela 9 (a) mostra os documentos por ano de publicação.

Referência (Ref.)	Metodologia e Método para construção de ontologias
R1	Tove
R2	Tove
R3	Tove
R4	Enterprise Ontology
R5	CYC
R6	Methontology
R7	Tove; Enterprise Ontology; CYC; Kactus
R8	Tove; Enterprise Ontology
R9	Tove
R10	Methontology; Enterprise Ontology; Tove
R11	Enterprise Ontology
R12	CYC; Tove; Enterprise Ontology
R13	CYC; Tove
R14	Tove; Enterprise Ontology; Methontology; KBSI IDEF5
R15	Enterprise Ontology; Tove; Kactus; Methontology; Sensus
R16	Methontology
R17	Enterprise Ontology; Tove; Methontology; Sensus; CYC
R18	Enterprise Ontology; Methontology
R19	Methontology; On-To-Knowledge
R20	On-To-Knowledge; Methontology; Enterprise Ontology; Método 101
R21	Tove, Methontology
R22	On-To-Knowledge ; CYC; Enterprise Ontology; Kactus; Methontology; Sensus; Tove
R23	Cyc; Enterprise Ontology; Tove; Kactus; Methontology; Sensus; On-To-Knowledge
R24	Cyc; Enterprise Ontology; Tove; Methontology; Kactus; Método 101
R25	On-To-Knowledge; Methontology; Enterprise Ontology; Cyc; Tove; Kactus; Sensus; Método 101

Tabela 10 – Mapeamento de metodologias encontradas na literatura

3.2.1.3 Passo 3 – Quantificação da ocorrência de citação

As metodologias e os métodos envolvidos foram referenciados por siglas de forma a facilitar a formalização da Tabela 11: método Enterprise Ontology (**Eo**); metodologia Tove (**Tv**); metodologia Methontology (**Mo**); método Kactus (**Kt**); método Sensus (**Su**); método Cyc (**Cyc**); metodologia On-To-Knowledge (**OtK**); método KBSI IDEF5 (**Kbs**); Método 101 (**101**).

A Tabela 11 apresenta a distribuição das metodologias e métodos por documento e a quantificação da ocorrência dos mesmos nos documentos selecionados.

Ref.	Eo	Tv	Mo	Kt	Su	Cyc	OtK	Kbs	101
R1		X							
R2		X							
R3		X							
R4	X								
R5						X			
R6			X						
R7	X	X		X		X			
R8	X	X							
R9		X							
R10	X	X	X						
R11	X								
R12	X	X				X			
R13		X				X			
R14	X	X	X					X	
R15	X	X	X	X	X				
R16			X						
R17	X	X	X		X	X			
R18	X		X						
R19			X				X		
R20	X		X				X		X
R21		X	X						
R22	X	X	X	X	X	X	X		
R23	X	X	X	X	X	X	X		
R24	X	X	X	X		X			X
R25	X	X	X	X	X	X	X		X
Total	**15**	**17**	**14**	**6**	**5**	**9**	**5**	**1**	**3**

Tabela 11 – Distribuição das metodologias por documento e freqüência de ocorrência

As metodologias e os métodos com maior freqüência de ocorrência foram a metodologia usada no **projeto Tove** (68%), o método proposto para o projeto **Enterprise Ontology** (60 %) e a metodologia **Methontology** (56%), sendo abordadas em mais de 50% dos trabalhos selecionados. Em seguida, vêm o método Cyc com 36 %, o método Kactus com 24 %, o método Sensus e a metodologia On-To-Knowledge com 20 %, o método 101 com 12% e o método KBSI IDEF5 com 4 %.

Para uma melhor visualização dos resultados, o gráfico da Figura 45 mostra as metodologias e os métodos que mais se destacaram nos documentos selecionados.

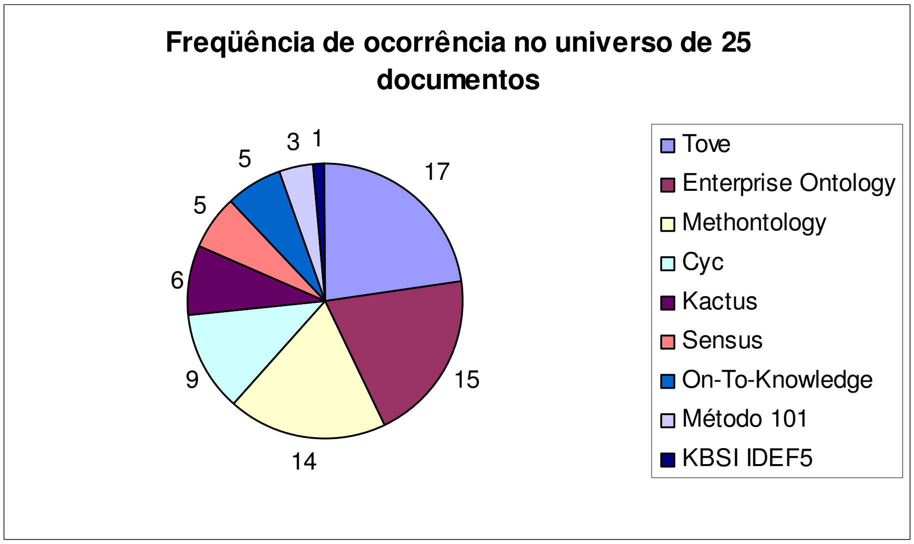


Figura 45 - Metodologias por ordem decrescente de freqüência de ocorrência

Na intenção de obter um conjunto razoável de metodologias e métodos para a análise comparativa, decidiu-se como **primeiro** critério de seleção: escolher as metodologias e métodos mais discutidos na literatura. Sendo assim, a tabela de freqüência de ocorrência foi utilizada como referência para a escolha. Dentre as metodologias candidatas à inserção no conjunto, o interesse da pesquisa se concentrava em metodologias que contemplassem a construção de **ontologias de domínio**, conforme explicitado no objetivo geral deste trabalho: *propor uma metodologia abrangente para a construção de ontologias para representação do conhecimento em sistemas de informações*. Ressaltando que, por metodologia abrangente entende-se uma metodologia que tenha a capacidade de cobrir um domínio de conhecimento. Diante disso, o **segundo** critério de seleção de metodologias foi investigar a ontologia resultante da aplicação de cada metodologia. Tal investigação é descrita na subseção a seguir.

3.2.1.4 Passo 4 – Determinação das metodologias e dos métodos

A investigação da ontologia resultante de cada metodologia foi feita mediante a análise de conteúdo nos documentos apresentados na Tabela 9 (a), bem como em outros identificados através de referências bibliográficas. No caso do projeto **Tove**, verificou-se que a metodologia foi derivada da experiência dos autores no desenvolvimento de ontologias para domínios corporativos (GRUNINGER e FOX, 1995). O método proposto por **Uschold** e **King**

foi baseado na prática da construção da ontologia de alto nível *Enterprise* (USCHOLD e KING, 1995). Ambos são capazes de construir ontologias de domínios sobre negócios. A metodologia **Methontology** já foi aplicada no desenvolvimento de uma ontologia no domínio da química (GÓMEZ-PEREZ, FERNANDEZ-LOPES e VICENTE, 1996). O método **Cyc** considera o conhecimento consensual sobre o mundo e pode ser utilizado na criação de ontologias para fundamentar diferentes sistemas inteligentes (LENAT, 1995; REED e LENAT, 2002). O método aplicado no projeto **Kactus** resultou numa ontologia no domínio de redes elétricas (BERNARAS, LARESGOITI e CORERA, 1996). O método baseado na ontologia **Sensus** foi aplicado no desenvolvimento de uma ontologia no domínio de planejamento de operação militar aérea (VALENTE et al, 1999). A metodologia **On-to-Knowledge** visa construir ontologias para aplicações de gestão do conhecimento e são altamente dependentes da aplicação (FERNANDEZ, GÓMEZ-PEREZ e CORCHO, 2004). O **método 101** já foi aplicado na construção de uma ontologia no domínio de vinhos (NOY e GUINNESS, 2001). E, finalmente, o método **KBSI IDEF5**, que auxilia a criação, a modificação e a manutenção de ontologias (JONES; BENCH-CAPON e VISSER, 1998).

Mediante os critérios estabelecidos, apenas dois objetos, dentre os nove apresentados, não foram selecionados para análise comparativa nesta pesquisa. A metodologia On-to-Knowledge depende altamente da aplicação na qual a ontologia será modelada, mostrando-se, portanto, inviável na construção de ontologias de domínio. Além disso, não foram identificadas na pesquisa ontologias resultantes da aplicação do método KBSI IDEF5**.** Sendo assim, as metodologias e os métodos escolhidos para a atividade de análise comparativa foram: a) metodologia de **Gruninger e Fox**; b) método de **Uschold e King**; c) metodologia **Methontology**; d) método **Cyc**; e) método **Kactus**; f) método **Sensus**; e g) **método 101**.

Acreditou-se que, analisando as metodologias mais discutidas na literatura poder-se-ia obter um resultado razoável no que diz respeito a uma proposta metodológica para construção de ontologias.

3.2.2 Seleção das metodologias para construção de vocabulários controlados

Na atividade de identificação de metodologias para construção de vocabulários controlados, foi realizada uma investigação na literatura da área de Ciência da Informação, visto ser um tema de propriedade desta área. Na investigação, constatou-se a existência de normas para construção de tesauros, taxonomias, enfim, vocabulários controlados, criadas por

entidades com aceitação internacional, propiciando, aos desenvolvedores, subsídios metodológicos para melhor orientá-los na confecção destes instrumentos. Afora as normas, guias nacionais também têm sido criados para ajudar na elaboração de vocabulários controlados. Algumas dessas normas e guias são descritas a seguir:

- Normas e diretrizes da **Unesco** de 1973: *Guidelines for the establishment and development of monolingual thesauri.*
- Manual de Elaboração de Tesauros Monolíngues de 1990: publicado em Brasília pelo Programa Nacional de Bibliotecas das Instituições de Ensino Superior.
- A versão em português da ISO 2788: Diretrizes para o estabelecimento e desenvolvimento de Tesauros Monolíngües, publicada pelo **Senai/IBICT** em 1993.
- Norma elaborada pela *American National Standards Institute National Information Standards Organization* (U.S.): *Guidelines for the Construction*, *Format*, *and Management of Monolingual Controlled Vocabularies* de 2005 ou ANSI/NISO Z39.19-2005.

Nessa pesquisa, foi utilizada como referência uma norma que se justificou pelos seguintes critérios: a) norma mais atual (de 2005), aceita e consolidada na comunidade de Ciência da Informação; e b) norma que apresenta abordagem interdisciplinar das teorias advindas da Ciência da Informação e da Terminologia - Teoria da Classificação Facetada; Teoria do Conceito; e Teoria da Terminologia - na construção de vocabulários controlados. Esta norma é a **ANSI/NISO Z39.19-2005**, que foi construída tendo como base as diversas normas americanas e internacionais sobre a criação de tesauros, incluindo a ISO 2788. Deu-se maior ênfase à norma ANSI/NISO Z39. 19-2005 por considerá-la mais atual em relação à norma ANSI/NISO Z39. 19-2003.

Destaca-se que no período de execução da atividade de coleta e análise dos dados, ficou constatado que a norma ANSI/NISO Z39. 19-2005 não seria suficiente para classificar seus conteúdos nas categorias de análise fundamentadas na norma IEEE-1074 (1997), conforme poderá ser conferido na seção 4.8. Tal evidência resultou na seleção do manual disponível no sítio da *Biblioteconomia, Informação e Tecnologia da Informação* – BITI (CAMPOS, GOMES e MOTTA, 2004) sobre elaboração de tesauros. É válido ressaltar que a metodologia explicitada no manual da BITI é conseqüência de trabalhos da professora Hagar Espanha Gomes e equipe (GOMES et al., 1990), os quais se mostram como referência para

cursos de Biblioteconomia e Ciência da Informação, além de Bibliotecas preocupadas com o desenvolvimento de terminologias adequadas para a representação do conteúdo de seus acervos.

Desse modo, os objetos selecionados como metodologias para construção de vocabulários controlados foram o manual da BITI e a norma ANSI/NISO Z39. 19-2005.

3.3 Coleta e análise dos dados

Esta seção descreve como foi conduzida a atividade de coleta e análise do material empírico. O primeiro passo desta etapa foi definir as categorias de análise que pudessem subsidiar a análise de conteúdo dos materiais envolvendo as metodologias e os métodos para construção de ontologias e a metodologia e a norma para construção de vocabulários controlados.

A maioria dos procedimentos de análise qualitativa organiza-se em torno de categorias (BARDIN. 1977). Na análise de conteúdo, as categorias são classes que reúnem um grupo de elementos (unidades de registro) em razão de características comuns. Segundo Bardin (1977), a escolha de categorias pode envolver vários critérios: i) **semântico** (temas); ii) **sintático** (verbos, adjetivos, pronomes); iii) **léxico** (juntar pelo sentido das palavras, agrupar os sinônimos, os antônimos); e iv) **expressivo** (agrupar as perturbações da linguagem, da escrita). De acordo com o autor, as categorias devem possuir certas características, a saber: i) **exclusão mútua** – cada elemento só pode existir em uma categoria; ii) **homogeneidade** – para definir uma categoria é preciso haver só uma dimensão na análise; iii) **pertinência** – as categorias devem dizer respeito às intenções do investigador, aos objetivos da pesquisa às questões norteadoras, às características da mensagem, etc.; iv) **objetividade e fidelidade** – se as categorias forem bem definidas, não haverá distorções devido à subjetividade dos analistas; e v) **produtividade** – as categorias serão produtivas se os resultados forem férteis em inferências, em hipóteses novas, em dados exatos.

Na análise de conteúdo feita nos documentos, descrita na seção 3.2, percebeu-se certa semelhança entre algumas fases de desenvolvimento dos instrumentos (ontologias e vocabulários controlados) e outras advindas do processo de desenvolvimento de *software*, conforme foi fundamentado na seção 2.3.3.5.1. Algumas dessas semelhanças foram identificadas principalmente nas atividades de análise de domínio, e nas abordagens técnicas para criação de modelos conceituais. Considerando que as ontologias de sistemas de informações são desenvolvidas segundo um processo de modelagem de domínio e

codificação, e dada a falta de uma estrutura de trabalho mais específica para análise de tais objetos, decidiu-se, na perspectiva deste trabalho, utilizar como subsídio para definição das categorias de análise o padrão aceito internacionalmente para desenvolvimento de *software*, a norma IEEE-1074 (1997).

O modelo de ciclo de vida de *software*, proposto na norma IEEE-1074, foi descrito na fundamentação teórica dessa pesquisa, e aqui é considerado um padrão aceitável para análise e comparação das fases de desenvolvimento de ontologias e vocabulários controlados. A justificativa para tal consideração é de que o modelo de ciclo de vida descreve o processo estruturado e metódico de desenvolvimento (PRESSMAN, 2002), e advém da "Engenharia de Software, uma disciplina considerada madura no sentido de possuir metodologias amplamente aceitas"[109] (FERNANDEZ, 1999, p.1). E, como as ontologias podem ser consideradas componentes de produtos de *software*, elas podem ser desenvolvidas de acordo com padrões usualmente indicados para *software* (FERNANDEZ, 1999). Tais padrões devem ser adaptados de acordo com o processo de construção de ontologias, conforme foi explanado na seção 2.3.3.5.2: i) planejamento; ii) especificação de requisitos; iii) aquisição de conhecimento; iv) conceitualização; v) formalização; vi) integração; vii) implementação; viii) avaliação; e ix) documentação.

Fernandez (1999, p.3) estabelece critérios para analisar metodologias para construção de ontologias. Alguns desses critérios foram utilizados para julgar a maturidade da metodologia frente à norma IEEE-1074 e são enumerados como segue: i) processos moldados no ciclo de vida de um *software*; ii) processo de gerenciamento de projeto; iii) processos de desenvolvimento orientado: pré-desenvolvimento, desenvolvimento e pós-desenvolvimento; e iv) processo integral. O critério **i** considera que uma ontologia deveria ser construída a partir das etapas do ciclo de vida de um *software*, ou seja, desde o planejamento até os testes de *software*; o critério **ii** busca a garantia da qualidade do produto a partir da gestão do projeto; em **iii** leva-se em consideração que o projeto passe por um estudo de ambiente, estudo de viabilidade, especificação de requisitos, conceitualização, formalização, implementação, avaliação, validação, manutenção e acompanhamento do uso pós-desenvolvimento; e, finalmente, o critério **iv** enfatiza a importância do treinamento dos usuários que ficarão responsáveis pela manutenção da ontologia, bem como o desenvolvimento de documentação em todo o ciclo de vida.

---

[109] "Software Engineering […] can be said to have reached adulthood, because it has widely accepted methodologies."

Finalmente, as categorias de análise de conteúdo do material empírico foram definidas a partir dos princípios elucidados por Bardin (1977) anteriormente. O critério de escolha das categorias foi o semântico, ou seja, de acordo com a norma IEEE-1074 (1997) e com a literatura da área de ontologias. Tais categorias foram, então, adaptadas diante aos processos extraídos da norma e características particulares às ontologias. São elas: i) **gerenciamento do projeto**; ii) **pré-desenvolvimento**; iii) **especificação de requisitos**; iv) **modelagem conceitual**; v) **formalização**; vi) **implementação**; vii) **manutenção**; viii) **integração**; ix) **avaliação**; e x) **documentação**.

Ressalta-se que a categoria *formalização* foi direcionada para o contexto das ontologias, que especificam modelos semi-computáveis através de uma linguagem formal, como também para o contexto dos vocabulários controlados, quando se define as suas formas de apresentação.

A definição das categorias de análise pode ser justificada a partir de uma visão genérica da engenharia de software, que abrange a análise, o projeto, a construção, a verificação e a gestão de elementos técnicos ou sociais (PRESSMAN, 2002). Desse modo, todas as categorias definidas nessa pesquisa se enquadraram em três fases genéricas, as quais, segundo Pressman (2002), independem da área da aplicação, do tamanho do projeto ou de sua complexidade: i) a *fase de definição*, que se concentra no *quê*, isto é, que comportamento é esperado do sistema; ii) *fase de desenvolvimento*, que focaliza o *como*, isto é, durante o desenvolvimento deve-se definir como os dados devem ser organizados (no caso das ontologias e dos vocabulários controlados, seriam os conceitos), implementados e testados; e iii) *fase de manutenção*, que se ocupa das modificações associadas com correções de erros, adaptações necessárias e melhoramentos frente aos requisitos levantados.

Definidas as categorias de análise, o próximo passo foi a elaboração dos instrumentos para coleta e registro dos dados. Para cada metodologia, método e norma envolvida foi elaborada uma tabela contendo um espaço dedicado a cada categoria de análise, conforme podem ser vistas no capítulo 4. Desta forma, tornou-se possível colher o conteúdo nos documentos envolvidos (relacionados em anexo) a partir das categorias e registrá-lo formalmente em cada tabela. Se uma categoria não fosse pertinente a alguma metodologia ou método, a coluna era preenchida com o valor "Ausente". A coleta dos dados passou pelas atividades de compreensão, análise e síntese de conteúdo, realizadas a partir do conhecimento adquirido na pesquisa. Tais atividades foram necessárias para a realização da etapa seguinte, que consistia em analisar comparativamente as categorias para cada metodologia, método e

norma; sob a orientação dos parâmetros estabelecidos para esta análise. Tal etapa é descrita na seção seguinte.

3.4 Análise comparativa das metodologias, dos métodos e da norma

Feita a coleta e registro dos dados, o próximo passo foi o desenvolvimento de um escopo comparativo das fases de desenvolvimento apresentadas nas metodologias e métodos para construção de ontologias e vocabulários controlados, segundo atividades advindas da norma internacional IEEE-1074 e atividades específicas às ontologias, como formalização e integração. Tal escopo foi concebido através de duas estruturas matriciais representando em suas colunas os objetos investigados e em suas linhas cada fase do ciclo de vida. A proposta da primeira matriz foi de apresentar o nível de detalhe acerca de cada metodologia, método e norma investigada. Já a segunda se propôs apresentar sinteticamente conteúdos que seriam de interesse à etapa final da pesquisa: *proposta metodológica para construção de ontologias*. A partir do tratamento e organização do conteúdo em tais estruturas, foi possível analisar comparativamente cada metodologia, método e norma e tirar conclusões teóricas e empíricas acerca do processo de construção de ontologias.

A partir da análise comparativa dos métodos, da norma e das metodologias estudadas, obteve-se um arcabouço teórico-metodológico que facilitou a proposta de uma metodologia para construção de ontologias fundamentada nas teorias da Ciência da Informação, da Computação, bem como em padrões reconhecidos internacionalmente para concepção de *software*. A próxima seção descreve como foi conduzida a atividade de proposição da metodologia para construção de ontologias.

3.5 Proposição da metodologia para construção de ontologias

A proposta metodológica de construção de ontologias foi desenvolvida no período de abril a maio de 2008, a partir da análise de duas metodologias e cinco métodos para construção de ontologias, além de uma norma e uma metodologia para construção de vocabulários controlados.

A proposta buscou o preenchimento de lacunas existentes no âmbito da construção de ontologias, lacunas estas apresentadas no decorrer de toda a pesquisa. Resumidamente, pode-se apontar que: **a)** dentro do domínio de desenvolvimento de ontologias, as abordagens para a

sua construção, na maioria das vezes, são específicas e limitadas (FERNANDEZ, GÓMEZ-PEREZ e JURISTO, 1997; USCHOLD e GRUNINGER, 1996) e **b)** a conseqüência é a ausência de atividades padronizadas, o que propicia que os processos sejam conduzidos de forma artesanal e não com o rigor que se esperaria de uma atividade científica. Além disto, verifica-se a falta de explicações sistemáticas sobre como e onde serão usadas as abordagens teóricas dentro de seu processo de elaboração (JONES; BENCH-CAPON e VISSER, 1998; FALBO, 1998).

Diante disso, a metodologia foi proposta tendo-se um conjunto de métodos fundamentados da norma IEEE-1074 (1997), associado por métodos e técnicas interdisciplinares capazes de alcançar um objetivo de forma mais eficiente. A escolha dos princípios teóricos e metodológicos foi justificada mediante avaliação do escopo comparativo para construção de ontologias e vocabulários controlados, descrito na seção 3.4.

4 - Apresentação da Análise dos Objetos Empíricos

O presente capítulo apresenta a fase de coleta de dados proposta nos passos metodológicos descritos no capítulo 3. Os dados coletados foram obtidos a partir do material empírico selecionado na pesquisa, que contemplou as metodologias e os métodos para construção de ontologias, bem como a metodologia e a norma para construção de vocabulários controlados.

Na atividade de coleta dos dados foram utilizadas as categorias de análise definidas a partir de fundamentos descritos na seção 3.3, de forma a servirem de suporte para a classificação e a organização dos conteúdos envolvidos no material. Para a execução da classificação do conteúdo, foi necessário um entendimento claro sobre cada categoria de análise envolvida nas metodologias, métodos e norma investigados.

Após a coleta dos dados, tornou-se possível apresentar o resultado da análise das metodologias e dos métodos para construção de ontologias em ordem cronológica de evolução (1990 até 2001), e da metodologia e da norma para construção de vocabulários controlados frente ao padrão IEEE 1074 (1997) e alguns processos particulares às ontologias, como formalização e integração. Para cada metodologia, norma ou método foi apresentado o ciclo de atividades ou recomendações envolvendo os estágios de desenvolvimento do instrumento e, logo em seguida, uma tabela sintetizando a proposta de cada objeto investigado. A categoria de análise não explicitada na apresentação dos objetos se deu em função dos mesmos não apresentarem propostas das fases correspondentes a tais categorias em seus métodos ou metodologias de construção. Para tal, o valor "Ausente" seria direcionado à tabela sinóptica quando da ausência de alguma categoria.

A seção 4.1 apresenta a ontologia Cyc e o método utilizado para seu desenvolvimento. A seção 4.2 apresenta considerações sobre o projeto Tove e os procedimentos da metodologia de Gruninger e Fox. A seção 4.3 apresenta o projeto da *Enterprise Ontology* e abrange os procedimentos do método de Uschold e King. A seção 4.4 apresenta o método Kactus. A seção 4.5 apresenta os métodos e as técnicas da Methontology. A seção 4.6 apresenta a ontologia SENSUS e o método baseado em tal ontologia, o método Sensus. A seção 4.7 apresenta o método 101. E, finalmente, a seção 4.8 tece considerações sobre a norma para construção de vocabulários controlados e apresenta os procedimentos metodológicos envolvidos no manual da BITI para construção de tesauros.

4.1 Método Cyc

Nos anos de 1980, a *Microelectronics and Computer Technology* - MCC deu início a criação da Cyc[110], uma ampla base de conhecimento que considera o conhecimento consensual sobre o mundo, incluindo regras e heurísticas para dedução sobre objetos e eventos do cotidiano (CYC PROJECT, 2005). A linguagem de representação da Cyc é a CycL (REED e LENAT, 2002), considerada híbrida por combinar *frames* com cálculos de predicado. Tal linguagem possui uma máquina de inferência que permite herança múltipla, classificação automática, manutenção de *links* inversos, verificação de restrições, busca ordenada, detecção de contradição e módulo de resolução.

Várias aplicações têm integrado seus módulos à base de conhecimento Cyc e à máquina de inferência CycL (FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004), dentre as quais pode-se citar: a) sistema de integração de bases de dados heterogêneas, que mapeia vocabulário Cyc para esquemas de bases de dados; b) busca aprimorada de conhecimento sobre informação capturada; c) integração de terminologia estruturada, que permite usuários importar, gerenciar e integrar tesauros múltiplos; e) módulo de recuperação de informação da rede mundial de computadores, que usa ferramentas de processamento de linguagem natural para acessar a base de conhecimento Cyc e permitir estendê-la a partir de informações disponíveis na Web.

A Figura 46 ilustra os três processos que foram conduzidos no desenvolvimento da base de conhecimento Cyc em 1990 por Douglas Lenat e Ramanathan Guha (LENAT e GUHA, 1990[111] apud FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004), e as seções a seguir apresentam tais processos através das categorias de análise *especificação de requisitos*, *implementação* e *integração*.

---

[110] A Cyc é considerada uma ontologia pelo fato de ser usada como base para construir diferentes sistemas inteligentes que conseguem comunicar e interagir entre si (FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004).

[111] LENAT, D. B.; GUHA, R.V. *Building large knowledge-based systems*. Massachusetts: Addison-Wesley, 1990. 372p.

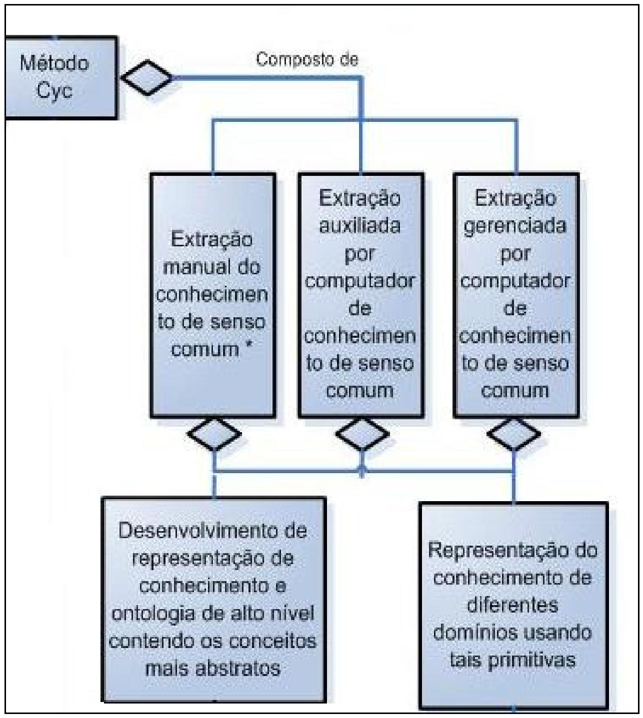


Figura 46 – Processos e atividades propostas pelo método Cyc
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.114)

4.1.1 Especificação de Requisitos e Implementação

No *primeiro* processo, o conhecimento requerido para a ontologia foi obtido de forma manual em diferentes fontes como artigos, livros e jornais, e o processo de aquisição de conhecimento ocorreu da seguinte forma:

- *Extração manual* do conhecimento requerido através da compreensão de artigos dispostos em livros e jornais.
- Investigação em artigos considerados extraordinários a fim de fazer uma *análise racional* dos motivos pelos quais tais artigos teriam sido classificados como extraordinários.
- *Identificação de questões* que “qualquer um” seria capaz de responder tendo apenas lido o texto. A base da Cyc seria ampliada a partir da capacidade de responder a tais questões.

Após a extração manual do conhecimento, o mesmo foi direcionado por meio de codificação (implementação através da linguagem CycL) à base de conhecimento Cyc.

O *segundo* processo foi conduzido de maneira automática, isto é, com uso de ferramentas computacionais de processamento de linguagem natural e aprendizado de

máquina capazes de usar conhecimento de senso comum suficiente para investigar e descobrir novos conhecimentos.

O *terceiro* processo foi conduzido por um número maior de ferramentas no sentido de gerenciarem a *extração de conhecimento de senso comum* (partes consideradas difíceis de serem interpretadas nas fontes de conhecimento envolvidas) na base Cyc.

O método Cyc vislumbra duas atividades, mostradas na Figura 46, que são executadas nos três processos elucidados acima: i) atividade de desenvolver uma ontologia de alto nível contendo conceitos mais abstratos; e ii) atividade de refinar a ontologia abstrata para representação do conhecimento desejado. Na *primeira* atividade deve-se identificar e codificar o conhecimento explícito e implícito existente nas fontes de conhecimento disponíveis, de forma que, na *segunda* atividade, novos conceitos e relações sejam inferidos com o auxílio de ferramentas de processamento de linguagem natural e aprendizado de máquina.

4.1.2 Integração

A base de conhecimento Cyc possui diferentes micro-teorias[112] com o propósito de reunir o conhecimento de diferentes domínios a depender do contexto (LENAT, 1995). Desse modo, ontologias de variadas complexidades têm sido mapeadas ou integradas com a Cyc, incluindo a SENSUS, uma ampla parcela da WordNet, dentre outras (REED e LENAT, 2002).

Reed e Lenat (2002) apresentam propostas baseadas em padrões tecnológicos de indústria para integrar sistemas computacionais à base Cyc, como DAML, XML *Schema* e UML. No caso da DAML, a linguagem pode ser importada e exportada na Cyc, tendo em vista que as propriedades da taxonomia da DAML são bem similares aos predicados em Cyc. Alguns subconjuntos de bases de conhecimento têm sido criados no formato DAML[113] . O mapeamento de XML *Schema* para Cyc é naturalmente mais difícil do que mapear uma ontologia DAML, porque o XML *Schema* geralmente não inclui informação taxonômica para classes e propriedades. Entretanto, os autores esperam que ferramentas que empregam análise semântica através de nomes de *tags* XML possam melhorar tal deficiência. E, finalmente, os autores destacam o mapeamento do núcleo de elementos da UML em Cyc, permitindo à base de conhecimento Cyc compreender melhor as interfaces de sistemas computacionais e seus

[112] Modelo formal do conhecimento a respeito de alguma área.
[113] http://www.daml.org

componentes, quando modelados na UML. Nesse sentido, os autores acreditam que a linguagem de modelagem unificada possa ser suficientemente expressiva[114] no sentido de permitir que a ontologia Cyc modele sistemas, como, por exemplo, um vocabulário em Cyc para modelar sistemas derivados da UML.

[114] A seção 2.3.2.2.1 apresentou algumas iniciativas de integração da UML com linguagens de representação de ontologias.

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Ausente	Simultânea à fase de implementação	Ausente	Ausente	**Extração manual e codificação** do conhecimento requerido. **Codificação** e **extração** do conhecimento com **apoio** de **ferramentas computacionais** como processamento de linguagem natural e aprendizado de máquina.	Ausente	**Padrões tecnológicos** de indústria têm sido propostos no sentido de integrar sistemas computacionais à **micro-teorias** da base Cyc, como os padrões DAML, XML *Schema* e UML.	Ausente	Ausente

Tabela 12 – Tabela sinóptica do método Cyc

4.2 Metodologia de Gruninger e Fox

A metodologia foi proposta por Michael Gruninger e Mark Fox em 1995 (GRUNINGER e FOX, 1995), tendo como base para o seu desenvolvimento a experiência obtida no projeto *Toronto Virtual Enterprise* – conhecido como projeto Tove (FOX, 1992; FOX et al., 1997), cujos princípios teóricos e metodológicos encontram-se na Inteligência Artificial.

O objetivo do projeto Tove é criar um modelo de senso comum sobre empresas, isto é, um conhecimento compartilhado sobre o negócio que conduza a deduções de respostas sobre questões acerca do domínio (FOX, 1992). Tais questões seriam a base para uma caracterização rigorosa acerca dos problemas que o modelo empresarial seria capaz de resolver (GRUNINGER e FOX, 1994a). Para tal, ontologias são criadas no sentido de especificar modelos para organizações públicas e privadas, levando em consideração as seguintes características: a) capacidade de fornecer uma terminologia compartilhada para organizações, que possa ser compreendida e utilizada por cada aplicação, isto é, para cada tipo de negócio; b) definição da semântica de cada termo através de uma teoria lógica; c) implementação da semântica em um conjunto de axiomas que permita à ontologia deduzir de forma automática respostas às questões comuns no escopo das organizações; d) definição de uma simbologia para representar graficamente termos ou conceitos (GRUNINGER e FOX, 1996).

A metodologia foi usada no *Enterprise Integration Laboratory* (Laboratório de Integração de Empresas) da *University of Toronto* (Universidade de Toronto) para o projeto e avaliação de ontologias integradas. A Figura 47 ilustra a metodologia.

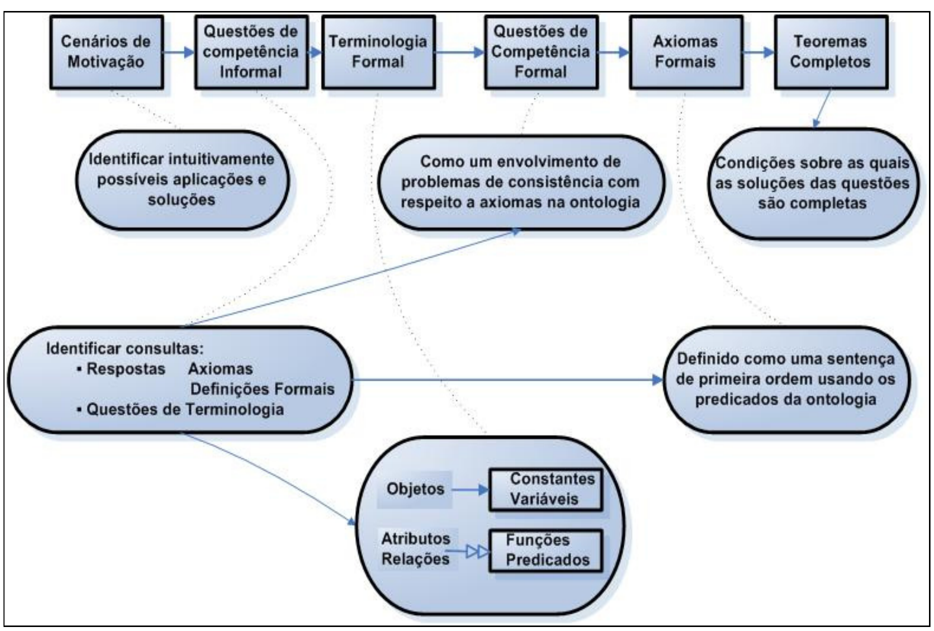


Figura 47 – Procedimentos propostos na metodologia de Gruninger e Fox
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.120)

A metodologia de Gruninger e Fox é considerada formal por utilizar declarações em lógica de primeira ordem na especificação de ontologias, obtendo vantagens acerca do poder de expressividade da lógica clássica (GRUNINGER e FOX, 1995). E pode ser usada como um guia para transformar cenários informais em modelos computáveis (FERNÁNDEZ-LOPEZ, GOMEZ-PEREZ e CORCHO, 2004, p.119).

As seções a seguir apresentam o ciclo de atividades da metodologia a partir dos procedimentos ilustrados na Figura 47, os quais possibilitaram o mapeamento para as seguintes categorias de análise: *pré-desenvolvimento*, *especificação de requisitos*, *modelagem conceitual*, *formalização*, *implementação*, *integração* e *avaliação*.

4.2.1 Pré-desenvolvimento

A fase pré-desenvolvimento da ontologia é iniciada pelos *cenários de motivação*, os quais descrevem problemas ou exemplos que não são cobertos adequadamente por ontologias existentes. A partir destes cenários-problema se chega a um conjunto de soluções possíveis que carregam a semântica informal dos objetos e relações que posteriormente serão incluídos na ontologia. Os cenários motivacionais serão importantes para a definição dos requisitos da ontologia, a qual deverá responder às questões de sua competência (GRUNINGER e FOX, 1995).

Segundo os autores da metodologia, qualquer proposta para uma nova ontologia ou extensão de uma ontologia deverá ser descrita nos cenários de motivação, bem como o conjunto de soluções pretendidas para os problemas apresentados em tais cenários. Desse modo, torna-se possível entender a motivação do propósito da ontologia em termos das aplicações que farão uso da mesma.

4.2.2 Especificação de Requisitos

Com base nos cenários de motivação, as *questões de competência informal* são elaboradas com a intenção de que seja possível representá-las e respondê-las usando a ontologia a ser desenvolvida. A formulação das questões é feita usando-se linguagem natural a fim de determinar o escopo da ontologia, as tarefas e os problemas que a ontologia pode solucionar antes de sua construção.

Segundo Gruninger e Fox (1995), as questões de competência não geram compromissos ontológicos. Elas são usadas para avaliar os compromissos ontológicos estabelecidos no domínio, isto é, verificar se a ontologia consegue responder às questões demandadas no domínio. Nesse sentido, tornar-se-ia possível avaliar a expressividade da ontologia (GRUNINGER e FOX, 1995). Um conjunto de problemas ou questões de competências serve para caracterizar as várias ontologias

e micro-teorias[115] no modelo de negócios. Gruninger e Fox (1995) apresentam duas ontologias para a modelagem de empresas, consideradas necessárias e suficientes para apoiar as tarefas em que são utilizadas: a *ontologia de atividade* e a *ontologia de organização*. Os autores exemplificam as seguintes questões de competência informal para a ontologia de atividade:

- Planejamento e programação: qual a seqüência de atividades deve ser preenchida para alcançar algumas metas? Quando elas devem ser iniciadas e terminadas?
- Projeção de tempo: dado um conjunto de ações que ocorrem em diferentes pontos do futuro, quais são as propriedades dos recursos e atividades que serão necessários?
- Acompanhamento da execução e de eventos externos: quais são os efeitos da ocorrência de eventos externos e inesperados sobre o modelo da empresa (interrupção de máquinas ou a indisponibilidade de recursos)?
- Raciocínio hipotético: o que vai acontecer se uma tarefa ocorrer antes do previsto e houver um atraso? Quais são os efeitos da compra de outra máquina sobre as encomendas?
- Competição baseada em tempo: planejar uma empresa que reduza o tempo do ciclo do produto, isto é, encontrar um plano de duração mínima que reduza as ocorrências nas ações e otimize as atividades de colaboração.

Já para a ontologia de organização[116] faz-se necessário definir como a mesma será integrada com a ontologia de atividades. As seguintes questões de competências informais são requeridas na ontologia de organização a fim de indicar o comportamento, a autoridade, a autorização, o compromisso e as metas dos agentes[117]:

- Quais atividades um agente particular deve executar?
- É possível para um agente executar uma atividade em alguma situação? Para tal, o agente tem habilidade para executar tal atividade?
- De forma a executar uma atividade em particular, quais permissões são necessárias?
- O agente permitiu executar uma atividade em alguma situação?
- Quais metas um agente se comprometeu a alcançar?
- Quais restrições de autoridade são necessárias entre um conjunto de agentes de forma a alcançar uma meta?
- Quais metas não são possíveis de serem realizadas por um agente?

[115] As micro-teorias contêm um conjunto de axiomas necessários e suficientes para representar e solucionar questões, pois fornecem uma semântica declarativa ao sistema (GRUNINGER e FOX, 1994a).
[116] Conjunto de restrições nas atividades executadas por agentes. A projeção de tempo, por exemplo, é usada para caracterizar as restrições que os agentes devem satisfazer de forma a serem capazes de executar as atividades (GRUNINGER e FOX, 1995).
[117] Qualquer entidade capaz de executar uma ação dentro da empresa.

- Para executar uma meta é necessária a assistência de outros agentes? Se sim, quais agentes são necessários?

4.2.3 Modelagem Conceitual

Apesar da concepção da estrutura conceitual das ontologias do projeto Tove ser realizada formalmente através de lógica de primeira ordem, achou-se pertinente classificar tal atividade na fase de modelagem conceitual e não na fase de formalização, visto que modelagem é uma atividade de apreensão do conhecimento sobre o domínio. No caso de construção de ontologias, o processo de elaboração da conceitualização é conduzido em tal fase, conforme será visto adiante nessa seção.

Através das questões de competência informal, o *ontologista* consegue extrair a *terminologia* que irá ser *formalmente* representada através do significado dos conceitos, atributos e relações na linguagem de lógica de primeira ordem ou uma linguagem equivalente como o KIF[118]. Através das respostas em linguagem natural para as questões, o *ontologista* extrai o conhecimento de forma a ser incluído nas definições formais dos conceitos, relações e axiomas. Dessa forma, a ontologia fornece a linguagem que será usada para expressar as definições e as restrições nos axiomas, além de propiciar a terminologia necessária ao tratamento das questões de competência informal (GRUNINGER e FOX, 1995, p.5).

Gruninger e Fox (1994a) destacam que as entidades no modelo Tove são representadas por objetos que especificam propriedades e relações. Tais objetos são estruturados em taxonomias, cujos passos para criação são elucidados por Gruninger e Fox (1995): i) identificar os **objetos** no domínio de discurso, os quais serão representados por constantes e variáveis no idioma escolhido; e ii) identificar as propriedades dos objetos, isto é, os **predicados** unários e binários. Os predicados unários são usados para representar conceitos; e os predicados binários (ou n-ários) usados para representar atributos e relações binárias (ou n-árias). Fernández-Lopez, Gomez-Perez e Corcho (2004, p.122) fornecem exemplos de predicados especificados em lógica de primeira ordem: a) representando conceito: *transport-means* ($transport); b) representando atributo: *traveler-name* ($traveler, $string); e c) representando relação binária: *has-destination* ($travelinfo,$location).

Para o desenvolvimento da terminologia dentro do projeto Tove, os cálculos situacionais[119] (*situation calculus*, em inglês) têm sido adotados (GRUNINGER e FOX, 1995; GRUNINGER e FOX, 1996; FOX et al., 1997) para fornecer semântica às ontologias envolvidas da modelagem de empresas. A idéia que envolve os cálculos situacionais é que existe uma situação (estado) inicial e

[118] *Knowledge Interchange Format* (KIF) é uma linguagem projetada para utilização no intercâmbio de conhecimentos entre os diferentes sistemas de computador.
[119] Informações adicionais sobre a formalização lógica dos cálculos situacionais podem ser encontradas em Gruninger e Fox (1996).

que mudanças de estados podem acontecer quando da execução de certas ações. As situações possíveis são representadas por uma estrutura em árvore, na qual duas seqüências diferentes de ações resultam em situações diferentes. Desta forma, cada nó que inicia uma situação pode ser entendido como um futuro de ações hipotéticas. A estrutura da árvore de cálculos situacionais apresenta todos os possíveis caminhos a partir dos quais os eventos no mundo podem desenvolver-se. Então, qualquer seqüência arbitrária de ações identifica um nó na árvore de situações (GRUNINGER e FOX, 1995).

4.2.4 Formalização

A fase de formalização da metodologia de Gruninger e Fox pode ser dividida em dois passos: i) formalização das questões de competência a partir da terminologia formal concebida na fase de modelagem conceitual e ii) especificação dos axiomas formais para representar e resolver as questões de competência, proporcionando, deste modo, semânticas declarativas à ontologia. Tais passos são apresentados a seguir.

4.2.4.1 Formalização das questões de competência

Após o levantamento das questões de competência informal na fase de especificação de requisitos e definição da terminologia formal ou conceitualização na fase de modelagem, as *questões de competência* são definidas *formalmente* como um vínculo de problemas consistentes em relação aos axiomas na ontologia (GRUNINGER e FOX, 1995), os quais serão definidos no passo especificado na seção 4.2.4.2. Para tal, criam-se regras descritas em linguagem formal, a fim de definir semanticamente as sentenças envolvidas nas questões.

Gruninger e Fox (1995, p.6) ressaltam que "todos os termos nas sentenças das questões de competência formal devem ser incluídos na terminologia da ontologia"[120] , de forma a estabelecer um vínculo entre a ontologia e as questões que ela precisará resolver. Segundo os autores, uma ontologia pode ser considerada distinta por sua capacidade de representar e resolver um conjunto de diferentes questões de competência em relação a outras ontologias.

4.2.4.2 Especificação dos axiomas em lógica de primeira ordem

Uma outra atividade que cabe à fase de formalização é a declaração dos axiomas para especificar as definições semânticas dos termos na ontologia e restrições em suas interpretações. Os

[120] "[…] all terms in the statement of the formal competency questions must be included in the terminology of the ontology."

axiomas são definidos como sentenças em primeira ordem usando predicados da ontologia. Gruninger e Fox (1995) destacam que a especificação dos axiomas ainda não consiste na implementação da ontologia, e sim em uma especificação da mesma. Dois exemplos, adaptados de Fernandez, Gomez-Perez e Corcho (2004), apresentam axiomas em lógica de primeira ordem expressando que:

a) *o viajante é considerado jovem se e somente se a sua idade for menor ou igual a 29 anos.*

$\forall\$x$ (viajante($x) $\wedge$ ($\exists\$y$ integer($y) $\wedge$ idade($x,$y) $\wedge$ ($y < 30))
$\leftrightarrow$ viajante_jovem($x))

b) *o ônibus local é uma subclasse de ônibus*

$\forall\$x$ (onibus_local($x) $\rightarrow$ ônibus($x))

Segundo Gruninger e Fox (1995), o processo de definição dos axiomas é considerado o mais difícil na definição da ontologia. Entretanto, tal processo é guiado pelas questões previamente definidas de competência formal. Assim, os axiomas na ontologia devem ser necessários e suficientes para expressar as questões de competência e caracterizar as suas soluções. E, qualquer solução para a questão de competência deve estar vinculada com os axiomas da ontologia. A Figura 48 mostra tal vínculo entre as questões de competência e os axiomas e vice-versa.

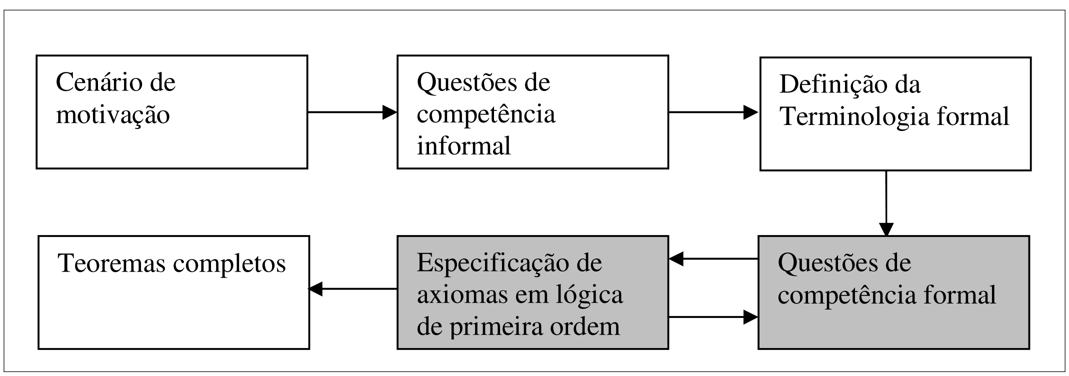


Figura 48 – Vínculo entre as formalizações das questões e dos axiomas
Fonte: Adaptado de Gruninger e Fox (1995, p.3)

Gruninger e Fox (1995) ressaltam que, se os axiomas propostos são insuficientes para representar as questões de competência formal e caracterizar as suas soluções, então é necessário adicionar objetos ou axiomas na ontologia até que sejam suficientes para a representação. O desenvolvimento de axiomas para a ontologia em relação às questões de competência é, portanto, um processo iterativo, segundo os autores.

Ainda segundo os autores, a expressividade de uma questão de competência está relacionada com o conjunto de axiomas que a representa. Pode-se comparar a expressividade de diferentes conjuntos de axiomas usando as questões de competência, isto é, se existe uma questão de competência a qual um conjunto de axiomas pode representar e outro não, então se considera que o

primeiro conjunto é mais expressivo do que o outro. Se dois conjuntos de axiomas podem representar uma questão de competência e caracterizar a sua solução, então tais conjuntos são equivalentes em relação à questão.

4.2.5 Implementação

Na metodologia proposta por Gruninger e Fox (1995), a fase de implementação não é descrita em detalhe. Segundo os autores, as especificações dos axiomas em lógica de primeira ordem são implementadas na linguagem Prolog-*Programming Logic*, que permitiria à ontologia deduzir de forma automática respostas às questões de competência. Fox et al. (1997, p.4) assinalam que o projeto Tove utiliza o poder do Prolog para conseguir provar os teoremas propostos.

4.2.6 Integração

Em algumas aplicações, pode existir uma ontologia de núcleo comum (*core ontologies*, em inglês) que é compartilhado com outras aplicações. Tal ontologia descreve genericamente entidades e conceitos de senso comum. Para tal, é necessário explicitar as características de relacionamentos entre o núcleo e as diferentes extensões, tendo em vista que diferentes comunidades usam extensões de ontologias específicas para suas aplicações.

O projeto Tove é composto por ontologias que representam um conhecimento comum de uma organização - *atividade*, *tempo*, *recurso*, *organização*, *produto*, *qualidade ISO-9000*, *custo baseado na atividade*, dentre outras consideradas importantes (GRUNINGER e FOX, 1994; GRUNINGER e FOX, 1994a; GRUNINGER e FOX, 1995; FOX et al., 1997; GRUNINGER e FOX, 1998). De forma a integrar tal conhecimento organizacional, os autores da metodologia indicam a formalização das diferentes perspectivas em relação a uma empresa para então integrá-las a ontologias de núcleo comum em "uma ferramenta de *software* que ajudará nas funções de engenharia de empresa [...]"[121] (GRUNINGER e FOX, 1994a, p.13). A Figura 49 mostra a integração de algumas ontologias no projeto Tove.

[121] "[…] a software tool that will support enterprise engineering functions […]"

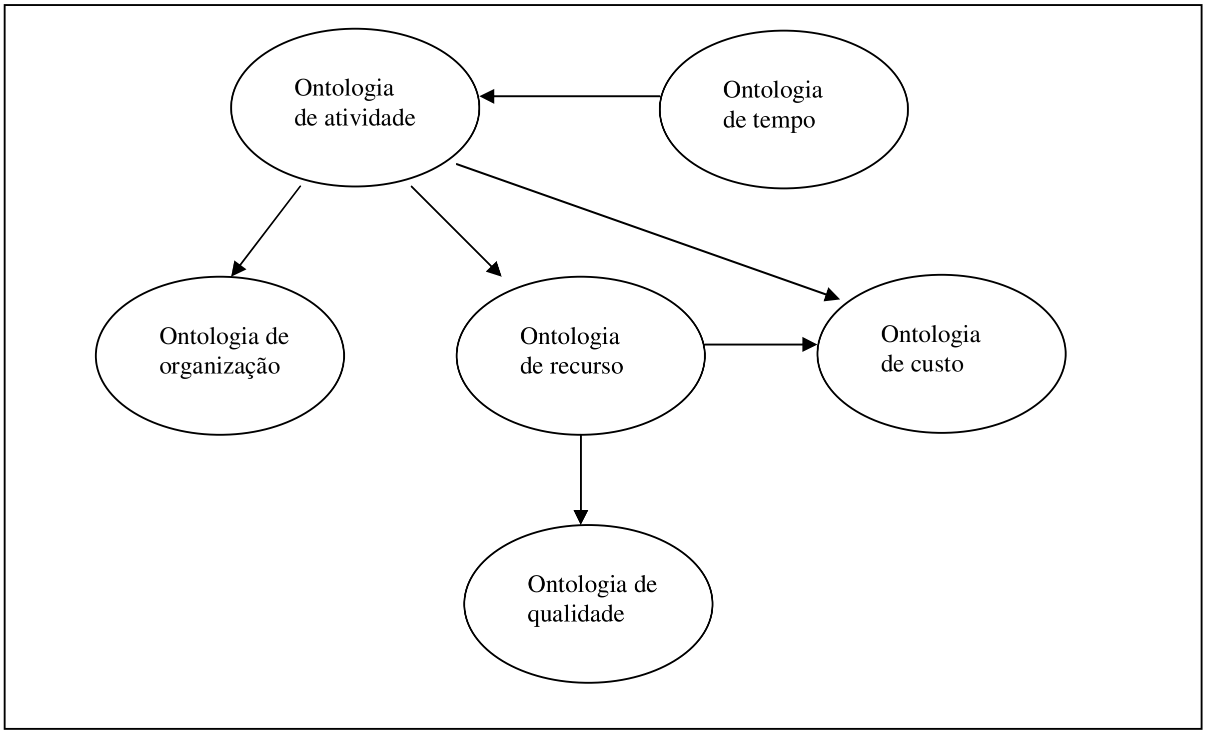


Figura 49 – Integração de algumas ontologias no modelo Tove.
Fonte: adaptado de Gruninger e Fox (1998, p.117)

Para tal, os axiomas responsáveis pela formalização do conhecimento em diferentes perspectivas devem ser suportados pelas ontologias de núcleo comum. De acordo com os autores da metodologia, "[...] a vantagem de especificar ontologias em lógica de primeira ordem é que somos capazes de representar e raciocinar sobre os compromissos ontológicos em relação a diferentes aplicações"[122] (GRUNINGER e FOX, 1995, p.7).

4.2.7 Avaliação

A fase de avaliação abrange uma série de testes em relação à competência da ontologia. Tal competência é definida por um conjunto de *questões formalmente descritas* elaboradas a partir dos estágios de especificação de requisitos, modelagem e formalização, conforme já elucidados na descrição dessa metodologia.

As definições das condições sobre as quais as soluções das questões são completas formam a base para os teoremas completos (*theorems completeness*, do inglês) da ontologia (GRUNINGER e FOX, 1995, p.9). Tais teoremas *verificam* se as soluções para as questões são completas e *determinam* a extensibilidade de uma ontologia através da explicitação do papel que cada axioma tem em provar um teorema. Assim, qualquer extensão da ontologia deve ser capaz de preservar os teoremas completos (GRUNINGER e FOX, 1995).

---

[122] "[...] the advantage of specifying ontologies in first-order logic is that we are able to represent and reason about the ontological commitments for different applications."

Kim, Fox e Gruninger (1999) apresentam um modelo de gerenciamento de qualidade ISO 9000, no qual é especificado pela ontologia denominada *Tove Quality Ontology-VB*. Segundo os autores, a representação da ontologia de qualidade pode ser usada para verificar se um requisito está em conformidade com o esperado. A ontologia de representação de requisitos fornece os termos, as relações e os atributos para representar os requisitos das características de um produto. Ela deve ser ainda capaz de responder às questões de competência usando os axiomas, os quais conseguem determinar se um dado requisito está em conformidade com a necessidade de qualidade.

Segundo Kim, Fox e Gruninger (1999), a *validação* da ontologia está na capacidade do sistema de informação, construído a partir da ontologia, em responder a questões de competência relativas a um domínio de conhecimento explorado.

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Descrição dos **cenários de motivação**.	Definição das **questões de competência informal**	Concepção da **terminologia** formal que servirá como **linguagem** para expressar as definições e as restrições dos axiomas.	Pode ser dividida em dois passos: i) **formalização das questões de competência** a partir da terminologia formal concebida na fase de modelagem conceitual e ii) **especificação dos axiomas formais** para representar e resolver as questões de competência.	As especificações dos axiomas em lógica de primeira ordem são implementadas na **linguagem Prolog**.	Ausente	Para **integrar** as ontologias envolvidas, a metodologia indica a **formalização** das **diferentes perspectivas** em relação a uma empresa para então integrá-las a **ontologias de núcleo comum**.	Os **teoremas completos** verificam se as soluções para as questões são completas.	Ausente

Tabela 13 – Tabela sinóptica da metodologia de Gruninger e Fox

4.3 Método de Uschold e King

O método foi proposto inicialmente por Mike Uschold e Martin King em 1995 (USCHOLD e KING, 1995) e estendido em 1996 por Mike Uschold e Michael Gruninger (USCHOLD, 1996; USCHOLD e GRUNINGER, 1996) na experiência de desenvolvimento da *Enterprise Ontology* (USCHOLD et al., 1998). Tal ontologia foi desenvolvida como parte do projeto *Enterprise* através do Instituto de aplicações em Inteligência Artificial da Universidade de Edinburgh e parceiros como IBM, Unilever e outros. A ontologia fornece uma coleção de termos e definições relevantes para as organizações, distribuídos em: a) classes de **alto nível** (relação, papel, ator, etc); b) classes sobre **atividades e processos** (atividades, recursos, planejamento, capacidades); c) classes sobre a **organização***,* (unidade, entidade legal, gestão, controle); d) classes sobre **estratégia** (propósito, estratégias, premissas), e e) classes sobre **marketing** (vendas, clientes, mercado).

Uschold e King (1995) consideram os seguintes estágios como sendo necessários a uma metodologia abrangente: 1) identificação do propósito e do escopo da ontologia; 2) construção da ontologia, que se divide em: i) captura; ii) codificação e iii) integração com ontologias existentes; 3) avaliação; e 4) documentação. A Figura 50 apresenta tais estágios envolvidos na construção de uma ontologia.

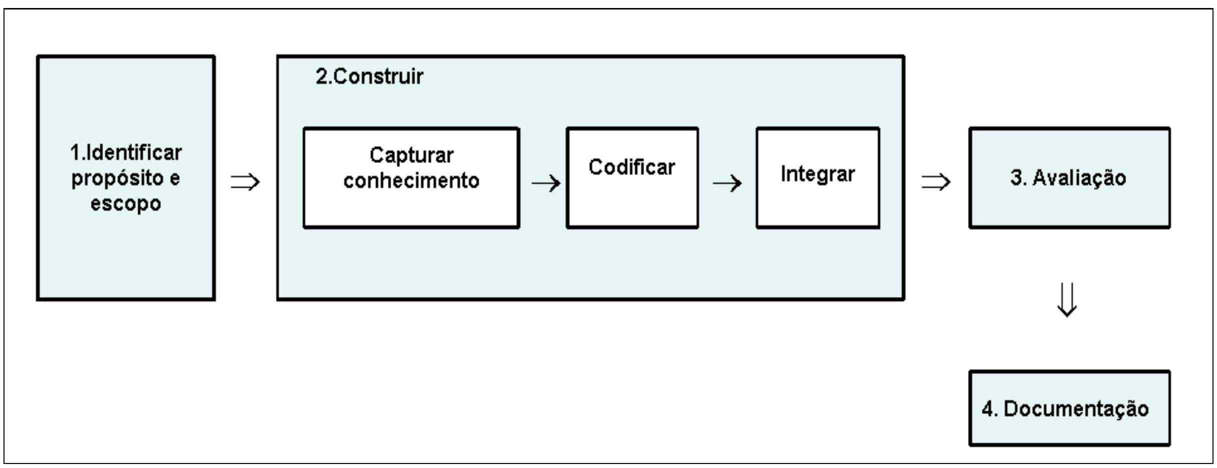


Figura 50 – Estágios do método de Uschold e King
Fonte: Adaptado de Uschold e King (1995)

Entretanto, na visão de Ushold e Gruninger (1996), os estágios não são suficientes para se definir uma metodologia, pois a mesma deveria incluir técnicas, métodos e princípios para cada um dos quatro estágios, além de indicar quais as relações existentes entre eles (ordem recomendada, intervalo, entradas e saídas).

As seções a seguir apresentam o ciclo de atividades do método de Uschold e King mapeado através das categorias de análise *especificação de requisitos*, *modelagem conceitual*, *implementação*, *integração*, *avaliação* e *documentação*.

4.3.1 Especificação de Requisitos

A especificação de requisitos é denominada por Uschold e King (1995) como um estágio de *identificação do propósito*, o qual objetiva identificar a necessidade de construção e as classes de usuários da ontologia, incluindo desenvolvedores, mantenedores e usuários das aplicações. Tal estágio também propõe conhecer o uso pretendido da ontologia como, por exemplo, reuso, compartilhamento, parte de uma base de conhecimento, dentre outros. Os autores do método também inserem as questões de competência como uma forma de elucidar o propósito da ontologia.

A proposta da metodologia unificada de Uschold (1996) considera três aspectos fundamentais ao processo de construção de ontologias, os quais se enquadram no estágio em pauta, a saber: a) formalismo; b) propósito da ontologia; e c) assunto da ontologia. O formalismo indica qual é o grau de formalismo pretendido sobre o vocabulário especificado, podendo ser de altamente informal, utilizando linguagem natural, até o rigorosamente formal, usando formalismos lógicos. O propósito da ontologia ressalta a intenção de uso da mesma e o assunto da ontologia contempla a natureza do assunto na qual a ontologia será caracterizada.

4.3.2 Modelagem Conceitual

Uschold e King (1995) recomendam a produção de um modelo do domínio antes da codificação ou implementação da ontologia. Para tal, denominam tal estágio de "captura da ontologia", porém não deixam claro se o mesmo pertence à fase de modelagem conceitual. Nessa pesquisa, "capturar a ontologia" se enquadraria na fase de modelagem conceitual, visto que tal fase se propõe a conceber um sistema organizado de termos, ou uma taxonomia, através da identificação dos conceitos, das relações e definições precisas acerca dos conceitos, relações e termos.

Alguns critérios para escolher um método para capturar a ontologia são propostos por Uschold e King (1995): i) cobertura do método quanto ao alcance de conceitos requeridos; ii) consideração do nível de detalhe e granularidade acerca dos conceitos do domínio pelo método; e iii) facilidade de aprendizado do método, ou seja, se existe uma documentação disponível.

No processo de modelagem conceitual, a atividade de **aquisição de conhecimento**, apresentada na seção 2.3.3.5.2, é realizada no sentido de: a) identificar os conceitos e as relações chave no domínio de interesse (também chamado de escopo); b) produzir definições textuais precisas e sem ambigüidades para os conceitos e relações; e c) identificar os termos com referência aos conceitos e relações de forma a conseguir um vocabulário consensual. Segundo Uschold e Gruninger (1996), as definições textuais não são criadas como um dicionário clássico, mas são construídas através da referência a outros termos, incluindo noções como classe e relação. Tais definições determinariam a representação do conhecimento na ontologia a ser usada.

Os autores do método recomendam a sessão de *brainstorming* como uma técnica para especificar termos e frases relevantes para o domínio, de forma a ajudar na elaboração do escopo da ontologia. Posteriormente seria produzido um documento em linguagem natural com a especificação de tais termos, de forma a ser útil na comunicação entre as pessoas participantes do processo de construção.

Na produção de definições para os conceitos e relações é importante considerar a existência de termos ambíguos. Podem ocorrer as seguintes situações: um termo corresponde a apenas uma definição, o que não causa problemas; vários termos correspondem a uma definição, indicando a existência de termos sinônimos, a partir dos quais é feito um registro perante a definição; e um termo corresponde a vários conceitos, o que configura ambigüidade. Para Uschold e Gruninger (1996), o tratamento de ambigüidades deve considerar: a) suspensão do uso do termo; b) esclarecimento da idéia relativa ao conceito, a partir de uma definição cuidadosa, e a utilização do menor número possível de termos técnicos; c) indicação dos conceitos importantes, justificando sua inserção na ontologia; e d) escolha de um único termo para o conceito.

O método de Uschold e King propõe três abordagens na identificação dos conceitos da ontologia, conforme elucidadas na seção 2.3.3.5.2: *bottom-up*, *top-down* e *middle-out*. Entretanto, Uschold e Gruninger (1996) recomendam o uso da abordagem *middle-out* por: "resultar em modelos estáveis e favorecer um controle sobre o nível de detalhes [...] reduzir a imprecisão [...] levando a menos re-trabalho e esforço"[123] (USCHOLD e GRUNINGER, 1996, p.21).

Na abordagem *bottom-up* primeiramente são identificados os conceitos mais específicos para então generalizá-los para conceitos mais abstratos. Algumas características desta abordagem seriam: i) resulta em um detalhamento de alto nível, aumentando o esforço geral; ii) torna-se difícil compatibilizar conceitos relacionados; e iii) Muitos dos termos podem não ser importantes na ontologia final, o que poderia levar a re-trabalho e ainda mais esforço.

A Figura 51 ilustra um exemplo da abordagem *bottom-up*, na qual os conceitos mais específicos (metrô de Londres, ônibus de Londres, táxi de Londres, metrô de Madri, ônibus de Madri, táxi de Madri) são identificados primeiramente e posteriormente agrupados (meios de transporte em Londres, metrô, ônibus local, táxi, meios de transporte em Madri) de acordo com as semelhanças identificadas entre eles.

---

[123] "[...] result in stable models, and keep the level of details in control [...] reduce inaccuracies […] leads to less re-work."

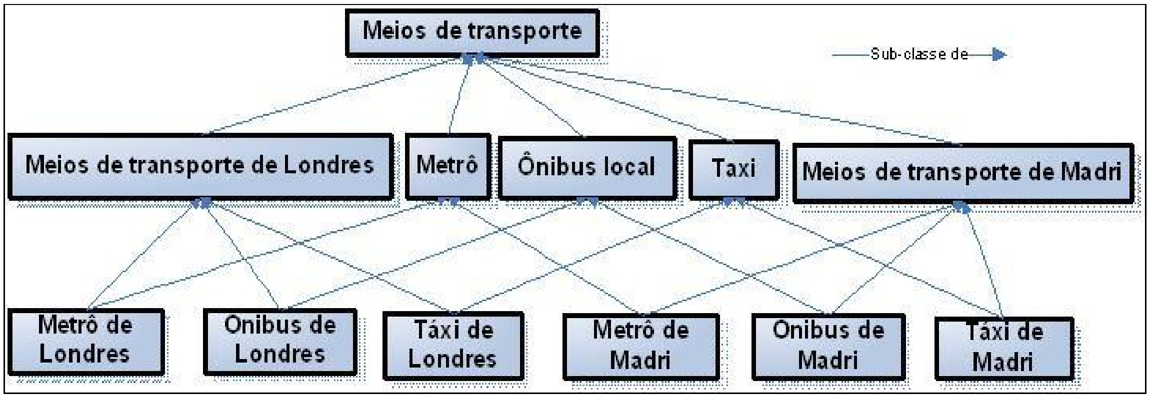


Figura 51- Exemplo de uma taxonomia construída a partir da abordagem *bottom-up*
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.117)

Na abordagem *top-down* os conceitos mais abstratos são identificados no primeiro momento, e então especializados para conceitos mais específicos. A Figura 52 mostra um exemplo de uma taxonomia construída utilizando a abordagem *top-down*. O conceito geral objeto é dividido em objeto concreto e objeto abstrato, e estes, por sua vez, são especializados de acordo com as características que os distinguem uns dos outros.

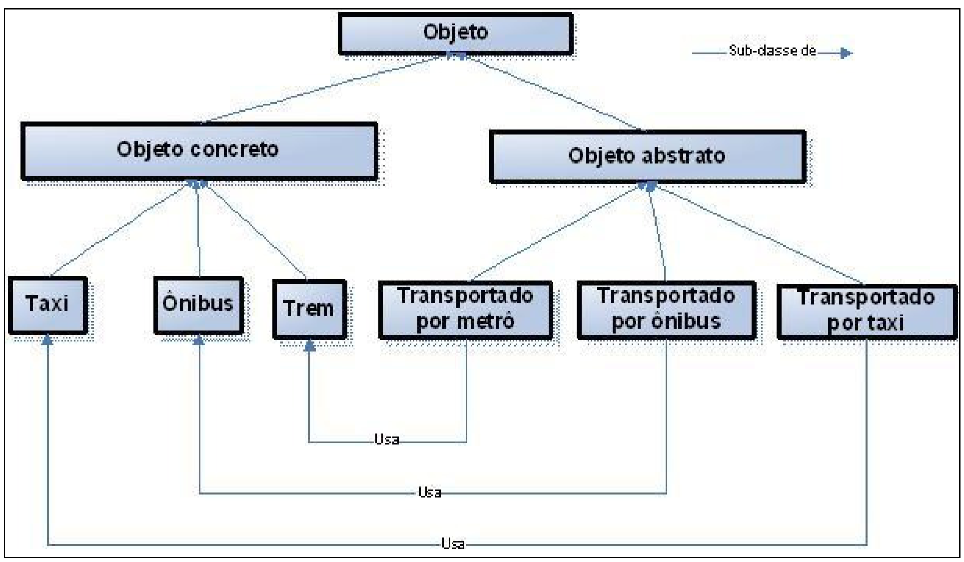


Figura 52 - Exemplo de uma taxonomia construída a partir da abordagem *top-down*
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.117)

A abordagem *top-down* teria como resultados: i) melhor controle do nível de detalhes; ii) escolha desnecessária de categorias de alto nível em função da imposição arbitrária; iii) menor estabilidade do modelo o que também leva a re-trabalho e mais esforço.

E, finalmente, a abordagem *middle-out*, que consiste em identificar primeiramente o núcleo dos conceitos básicos, para então especializá-los e generalizá-los na medida em que for necessário. A Figura 53 apresenta uma taxonomia, em que os conceitos básicos (metrô, ônibus e táxi) são

generalizados para a classe meios de transporte e o conceito ônibus é especializado para as subclasses ônibus local, microônibus e van.

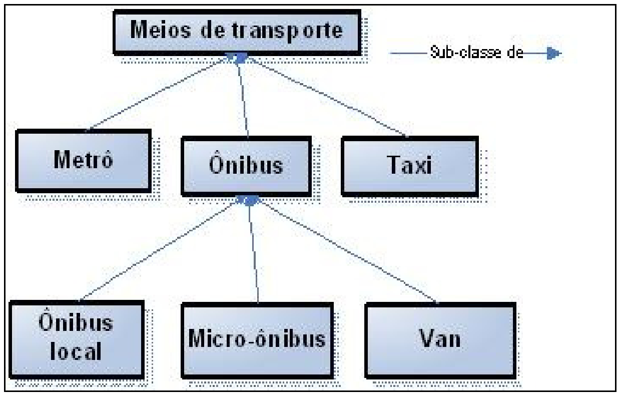


Figura 53 - Exemplo de uma taxonomia construída a partir da abordagem *middle-out*
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.118)

A abordagem possui as seguintes características: i) o detalhamento é considerado equilibrado; ii) os detalhes surgem quando necessários pela especialização dos conceitos básicos, de modo que alguns esforços sejam evitados; iii) as categorias de mais alto nível determinadas a partir de conceitos básicos crescem naturalmente e são mais estáveis. Isto implica em menos re-trabalho e menos esforço adicional.

4.3.3 Implementação

Esta fase objetiva representar explicitamente a conceitualização "capturada" nas fases anteriores: especificação de requisitos e modelagem conceitual. Tal representação requer a escolha de uma linguagem de representação formal de maneira a criar o código. Uschold e King (1995) citam algumas linguagens usadas para representar ontologias: Prolog, gráficos conceituais, L-Lilog, Ontolingua e linguagens da família KL-One (Back, Back++, LOOM, Classic). Os autores afirmam que as experiências do uso de tais linguagens podem servir de ponto de partida para o desenvolvimento de guias na escolha de linguagens de representação de ontologias.

Os autores definem critérios para a escolha da linguagem apropriada para representar ontologias. Nesse sentido, a linguagem deveria: i) apresentar clareza em seus elementos; ii) evitar distância conceitual, isto é, as primitivas semânticas da linguagem deveriam estar de acordo com a maneira de pensar dos usuários acerca dos conceitos passíveis de representação; iii) possuir poder de expressividade acerca dos conceitos envolvidos; iv) ser especificada frente a um padrão como, por exemplo, a ISO; v) ser traduzida e transportada para outras aplicações; vi) possuir suporte de algum método ou guia de implementação; vii) possuir semântica formal que possibilitaria a

verificação de consistências; viii) ser de fácil obtenção; e ix) ser flexível de modo a não exigir representações particulares.

Segundo Uschold e King (1995), a representação formal envolve um comprometimento ontológico com a meta-ontologia, isto é, com a ontologia de alto nível que se propôs integrar. A linguagem de implementação precisa ser coerente com os axiomas especificados na ontologia.

4.3.4 Integração

Durante os processos de captura e codificação da ontologia, existe uma questão de verificar a possibilidade de integração com ontologias existentes. Segundo Uschold e King (1995), o desenvolvimento de guias e ferramentas para integração pode ser considerado um grande desafio no desenvolvimento de uma metodologia abrangente. Um método indicado é explicitar todos os pressupostos subjacentes à ontologia de forma a alcançar uma correspondência semântica com a ontologia a qual se pretende integrar.

4.3.5 Avaliação

Segundo Uschold e King (1995), o procedimento de se avaliar ou julgar tecnicamente a ontologia, seus ambientes de *softwares* associados e a documentação pode ser feito através das especificações de requisitos, das questões de competência, bem como da realidade imposta pelo mundo.

4.3.6 Documentação

A documentação é definida de acordo com o propósito e tipo da ontologia. Segundo Uschold e King (1995), todas as pretensões da ontologia consideradas importantes devem ser documentadas, juntamente com os conceitos principais definidos na ontologia. Além das primitivas usadas para expressar as definições na meta-ontologia.

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Ausente	Identificar a **necessidade de construção, as classes de usuários** e o **uso pretendido** da ontologia.	a) **identificar** os **conceitos** e **relações** chave no domínio de interesse; b) **produzir definições textuais** precisas e sem ambigüidades para os conceitos e relações; e c) **identificar** os **termos** com referência aos conceitos e relações de forma a conseguir um vocabulário consensual.	Ausente	Objetiva representar explicitamente a conceitualização "capturada" nas fases anteriores. Tal representação **requer** uma **linguagem** de **representação** formal como Prolog e Ontolingua, de maneira a **criar** o **código**.	Ausente	**Verificar** a **possibilidade** de **integração** com **ontologias** já **existentes**. Um método indicado é **explicitar** todas as **pretensões** subjacentes à ontologia de forma a alcançar uma **correpondência semântica** com a ontologia a qual se pretende integrar.	Pode ser feita através das especificações de requisitos, **questões de competência**, bem como da realidade imposta pelo mundo.	A documentação é definida de acordo com o **propósito** e **tipo** da ontologia. Todas as **pretensões** consideradas **importantes** da ontologia devem ser documentadas.

Tabela 14 – Tabela sinóptica do método de Uschold e King

4.4 Método Kactus

A ênfase do projeto europeu *Esprit Kactus* está na organização de bases de conhecimento que podem ser compartilhadas e reusadas em diferentes sistemas baseados em conhecimento. Para tal, utiliza ontologias de domínio (SCHREIBER, WIELINGA e JANSWEIJER, 1995) para organizar o conhecimento independente da aplicação de *software* que será construída.

Baseando-se no projeto *Kactus*, Amaya Bernaras e colegas (BERNARAS, LARESGOITI e CORERA, 1996) investigaram a viabilidade da reutilização do conhecimento em sistemas de complexidade técnica, como o domínio de redes elétricas, e o papel das ontologias como suporte a tais sistemas. Tal investigação resultou em um método de construção de ontologias, cujos processos envolvidos estariam condicionados ao desenvolvimento da aplicação, ou seja, toda vez que uma aplicação fosse construída, a ontologia, que representa o conhecimento necessário para a aplicação, seria refinada.

As seções a seguir apresentam os processos propostos no método Kactus mapeados a partir das categorias *especificação de requisitos*, *modelagem conceitual* e *integração*.

4.4.1 Especificação de Requisitos

Essa fase contempla a especificação da aplicação que consiste em produzir um contexto para a mesma e uma visão dos componentes que serão necessários de serem modelados em tal aplicação.

O resultado da fase é uma *lista de necessidades ou requisitos* que precisam ser atendidos pela aplicação.

4.4.2 Modelagem Conceitual

Tendo desenvolvida a lista de requisitos, o próximo passo seria *identificar os termos relevantes* para o domínio da aplicação a partir de tais requisitos.

Posteriormente, em função dos termos identificados e das necessidades especificadas poder-se-ia *construir um modelo preliminar*, obtendo-se várias visões do modelo global em conformidade com categorias relevantes de alto-nível ontológico, por exemplo, conceito, relação, atributo, etc.

E, finalmente, o último passo contemplaria *refinar e estruturar a ontologia* a fim de obter um modelo definitivo. Para tal, tais atividades seriam realizadas de acordo com princípios de organização hierárquica e modularização (ontologias separadas para cada funcionalidade).

4.4.3 Integração

Para a construção do projeto preliminar na fase de modelagem, o método Kactus propõe buscar por ontologias já desenvolvidas por outras aplicações no sentido de reutilização das mesmas. As ontologias reutilizadas demandariam refinamento e extensão para serem usadas na nova aplicação. A Figura 54 mostra um exemplo de reutilização de ontologias de aplicações diferentes. A imagem de interseção precisaria ser refinada e estendida de acordo com as características do domínio da aplicação. Caso não seja possível a utilização de ontologias já existentes, o método indica a construção da ontologia a partir de esboço para posterior refinamento e estruturação da mesma.

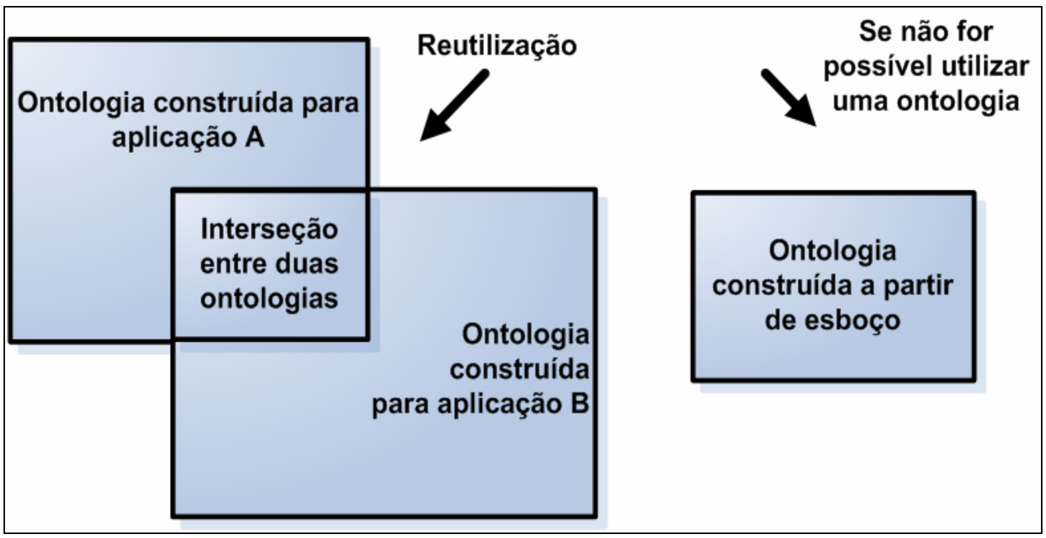


Figura 54 – Projeto do modelo de aplicação no método Kactus
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.124)

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Ausente	**Lista** de **necessidades** ou **requisitos** que precisam ser atendidos pela aplicação.	i) **identificar** os **termos relevantes** para o domínio da aplicação a partir dos requisitos; ii) **construir** um **modelo preliminar**, obtendo-se várias visões do modelo global; iii) **refinar** e **estruturar** a ontologia a fim de obter um modelo definitivo.	Ausente	Ausente	Ausente	**Buscar por ontologias já desenvolvidas por outras aplicações** no sentido de **reutilização** das mesmas. As ontologias reutilizadas demandariam refinamento e extensão para serem usadas na nova aplicação.	Ausente	Ausente

Tabela 15 – Tabela sinóptica do método Kactus

4.5 Metodologia Methontology

A metodologia para construção de ontologias *Methontology* foi desenvolvida no laboratório de Inteligência Artificial da Universidade Politécnica de Madri entre 1996 e 1997 pelo grupo de pesquisadores Mariano Fernández, Asunción Gómez-Pérez, Antônio J. de Vicente e Natalia Juristo (GOMEZ-PEREZ, FERNANDEZ e VICENTE, 1996; FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997). Projetos têm sido desenvolvidos em diferentes domínios utilizando-se de ferramentas como *Ontology Design Environment* - ODE (FERNANDEZ, et al., 1999) e WebODE (CORCHO et al., 2003), as quais fornecem suporte tecnológico à metodologia.

A Methontology contempla um conjunto de **estágios** de desenvolvimento, um **ciclo de vida** baseado em evolução de protótipos (explicado mais adiante nessa seção) e **técnicas** para realizar as **atividades** de planejamento, desenvolvimento e suporte. A atividade de planejamento inclui um escalonamento das tarefas e controle sobre as mesmas, no sentido de alcançarem a qualidade devida. As atividades de suporte contemplam aquisição de conhecimento, documentação e avaliação, e ocorrem durante todo o ciclo de vida da ontologia, conforme pode ser visualizado na Figura 55. Os estágios iniciais de desenvolvimento (especificação e conceitualização) implicam em um grande esforço dentro das atividades de suporte, como a aquisição de conhecimento e a avaliação. Várias são as razões: a) a maior parte do conhecimento é adquirida no início do processo de construção da ontologia; e b) deve-se avaliar corretamente o modelo conceitual para evitar futuros erros no ciclo de vida da ontologia. Por fim, a documentação detalhada deve ser produzida após cada estágio previsto no ciclo de vida.

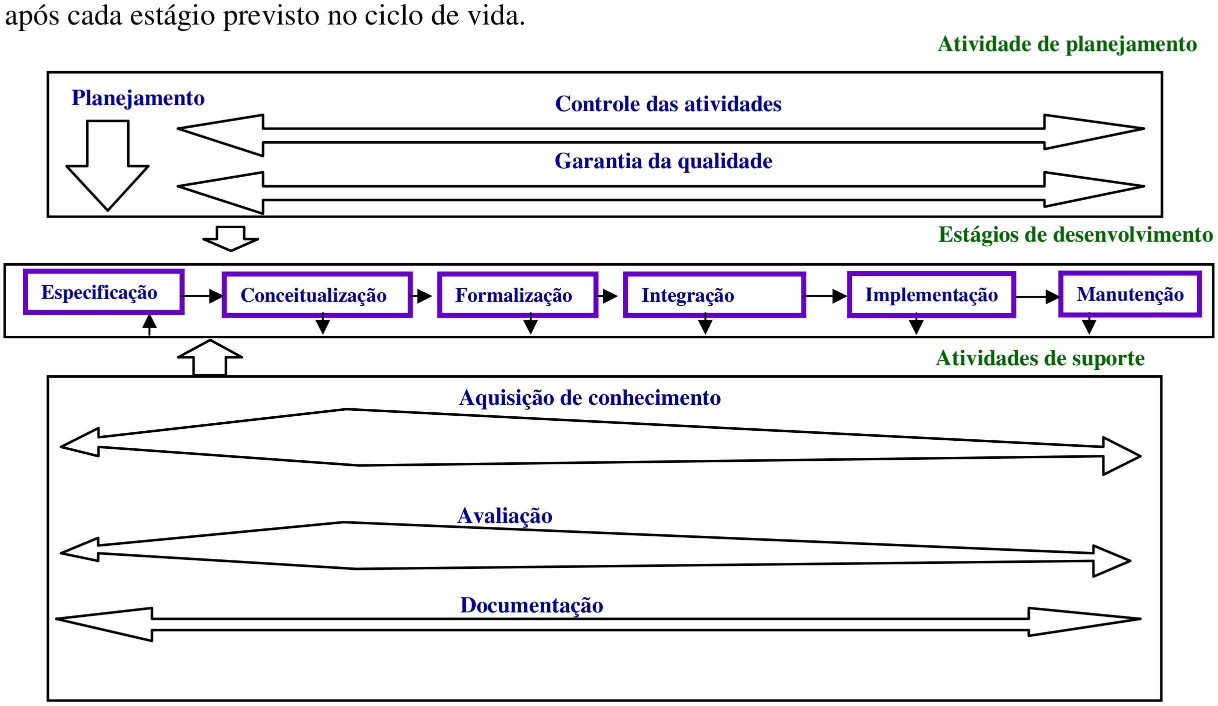


Figura 55 - Estágios e atividades do ciclo de vida da ontologia
Fonte: Adaptado de Fernandez, Gomez-Perez e Juristo (1997, p.35)

Os estágios e as atividades envolvidas no processo de construção de ontologias são considerados similares aos processos de desenvolvimento de *software* descritos na Engenharia de Software (modelos de ciclo de vida em cascata, incremental e evolutivo), bem como em técnicas advindas da Engenharia do Conhecimento, como a atividade de aquisição de conhecimento (FERNANDEZ, GOMEZ-PEREZ E JURISTO, 1997; FERNANDEZ, 1999; FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004). O ciclo de vida em **cascata** (ROYCE, 1987[124] apud FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997) é o modelo tradicional da Engenharia de Software, no qual as atividades são executadas em seqüência. Existe, porém, um problema nesta abordagem para as ontologias no sentido de não possibilitar inclusão, remoção ou modificação dos termos definidos na ontologia, pois não permite voltar aos estágios iniciais. O ciclo de vida **incremental** (McCRACKEN e JACKSON, 1982[125] apud FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997) resolve o problema de não se conseguir levantar todos os requisitos no estágio de especificação. Porém, tal abordagem implica em desenvolver a ontologia em camadas, permitindo a inclusão de novas definições somente quando uma nova versão for planejada. Já a **evolução de protótipos** permite a modificação, a inclusão e a remoção de definições conceituais na ontologia a qualquer momento a partir da necessidade de desenvolvimento de uma nova versão (protótipo). Os autores consideram essa última abordagem como a mais indicada para um ciclo de vida de construção de ontologias.

As seções seguintes descrevem em mais detalhe o ciclo de atividades da Methontology, baseado na evolução de protótipos, cujas categorias de análise utilizadas para tal descrição foram: *gerenciamento de projeto*, *especificação de requisitos*, *modelagem conceitual*, *formalização*, *implementação*, *manutenção*, *integração*, *avaliação* e *documentação*.

4.5.1 Gerenciamento do projeto

Para cada início de um protótipo, a Methontology propõe a execução do processo de gerenciamento, que contempla as atividades de: i) escalonamento; ii) controle e iii) garantia de qualidade. A atividade de escalonamento identifica as tarefas a serem executadas, a maneira como são organizadas e o período e os recursos necessários para a realização de cada uma delas. A atividade de controle garante que as tarefas escalonadas sejam completadas de modo adequado. E, finalmente, a atividade de garantia da qualidade busca assegurar que a ontologia, o *software* e a documentação sejam concluídos com a qualidade devida.

---

[124] ROYCE W. M. *Managing the Development of Large Software Systems*. Proc. 9th International Conference Software Engineering. IEEE. Computer Society. 1987. pgs.328-338.
[125] McCRACKEN; M A. JACKSON. *Life Cycle Concept Considered Harmful*. ACM Software Engineering notes. Abril 1982. Pags. 29-32.

O processo de gerenciamento de protótipos é executado em paralelo com as atividades de desenvolvimento durante todo o ciclo de vida da ontologia (FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004), conforme pôde ser visualizado na Figura 55.

4.5.2 Especificação de Requisitos

Esta fase tem o objetivo de produzir um documento de especificação da ontologia em linguagem natural, usando um conjunto de representações intermediárias (GOMEZ-PEREZ, FERNANDEZ e VICENTE, 1996; FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004) ou usando questões de competência (GRUNINGER e FOX, 1995). Tal especificação deve ser concisa (termos sem duplicidade ou irrelevantes), parcialmente completa (cobertura dos termos e nível de granularidade) e consistente (sentido dos termos e seus significados para o domínio). Pelo fato de utilizar um ciclo de vida baseado na evolução de protótipos, a Methontology possibilita retornar à etapa de especificação de requisitos (processo iterativo) a qualquer momento do projeto e fazer os devidos ajustes a depender da necessidade. Sendo assim, a especificação pode resultar, em princípio, num documento parcialmente completo.

De acordo com Fernandez, Gomez-Perez e Juristo (1997), as informações necessárias para a especificação de requisitos seriam: i) o propósito da ontologia; ii) o grau de formalidade - desde o informal até o formal; e iii) o escopo. O propósito da ontologia abrange a intenção de seu uso, os possíveis cenários que demandam o uso e os usuários que irão utilizá-la. O grau de formalidade vai depender da formalização envolvida na codificação dos termos e seus significados, podendo ser desde o informal até o rigorosamente formal, conforme classificado por Uschold e Gruninger (1996). E, finalmente, o escopo vai incluir um conjunto de termos, suas características e sua granularidade. Um exemplo de escopo pode ser conferido na Figura 56, na qual apresenta uma especificação de requisitos no domínio da química.

Documento de especificação da ontologia
Domínio: química
Data: 15 de maio de 1996
Conceituado por: Assunción Gómez-Pérez
Implementado por: Mariano Fernandez-Lópes
Propósito:
Ontologia sobre substâncias químicas, a ser utilizada para obtenção de informações sobre elementos químicos, com fins de ensino e análise. Essa ontologia pode ser utilizada para conferências, por exemplo, o peso atômico do elemento sódio.
Nível de formalidade: semi-formal
Escopo:
- Lista de 103 elementos de substâncias: lítio, sódio, cloro...
- Lista de conceitos: halogênios, gases nobres, semi-metais, metais...
- Ao menos informações sobre as seguintes propriedades: número atômico, peso atômico, volume atômico a 20 graus celsius, ponto de ebulição, densidade a 20 graus Celsius...
Fontes de conhecimento:
Manual de Química e Física, 65 Edição, CRC Press...

Figura 56 – Especificação de requisitos da ontologia no domínio da química
Fonte: Adaptado de Fernandez, Gomez-Perez e Juristo (1997, p.37)

A atividade de **aquisição de conhecimento** é conduzida na fase de especificação de requisitos no sentido de identificar os conceitos do domínio e elaborar o primeiro glossário de termos potencialmente relevantes que servirá à fase de modelagem conceitual. A aquisição de conhecimento é considerada uma atividade independente e, muitas das vezes, simultânea a outras atividades (especificação e modelagem conceitual) e diminui à medida que o desenvolvimento da ontologia avança, conforme discutido anteriormente e mostrado na Figura 55. As possíveis fontes de conhecimento são: especialistas do domínio, livros, manuais, figuras, tabelas, outras ontologias, dentre outras. E as possíveis técnicas para se obter conhecimento são: *brainstorming*, entrevistas estruturadas, análise informal e formal de documentos e ferramentas de aquisição de conhecimento.

Os autores da Methontology dão ênfase à estratégia de identificação de termos proposta por Uschold e Gruninger (1996), os quais argumentam que a abordagem *middle-out* é excelente para identificar os conceitos primários da ontologia e posteriormente especializá-los ou generalizá-los, tornando os conceitos mais estáveis, além exigir menos esforço e re-trabalho.

4.5.3 Modelagem Conceitual

A proposta da fase de modelagem conceitual é organizar e representar o conhecimento útil sobre um domínio. Para tal, a atividade de **aquisição de conhecimento** também é realizada nessa fase no sentido de refinar o glossário preliminar de termos desenvolvido na fase de especificação, e fornecer significados para os conceitos envolvidos. As técnicas de análise formal e informal de documentos em livros e manuais, em conjunto com entrevistas estruturadas e não estruturadas com especialistas, podem ser usadas para incluir ou remover termos no *glossário de termos*. Além disso, as entrevistas realizadas com especialistas podem ajudar na construção das árvores de classificações de conceitos (apresentadas adiante) de forma a confrontar com o que foi levantado nos documentos.

Após a construção do glossário de termos, o próximo passo seria agrupar os termos, como conceitos e como verbos. Em seqüência verificar as relações em comum entre os conceitos, bem como no caso dos verbos, de forma a direcioná-los para um conjunto apropriado de conceitos ou de verbos. Para tal, é importante identificar conjuntos de conceitos disjuntos, ou seja, conceitos que não possuem instâncias em comum. Finalmente, são construídas as *árvores de classificação de conceitos*, que organizam o domínio de conceitos em taxonomias e modularizam o domínio em ontologias independentes, e os *diagramas de verbos*, que organizam as relações entre os conceitos das taxonomias. Os autores da metodologia afirmam que após a construção da árvore de classificação de conceitos e do diagrama de verbos, o processo de desenvolvimento da ontologia pode ser dividido em diferentes equipes co-relacionadas. A Figura 57 apresenta de forma resumida as técnicas de representações intermediárias usadas na fase de modelagem conceitual para a concepção do modelo ou da conceitualização.

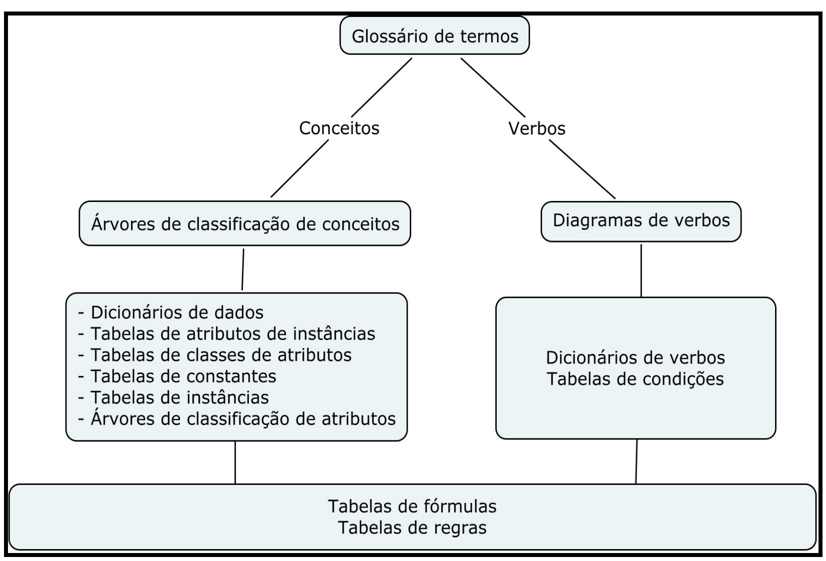


Figura 57 – Representações intermediárias na fase de conceitualização
Fonte: Adaptado de Fernandez, Gomez-Perez e Juristo (1997, p.38)

Para a construção da árvore de classificação de conceitos, a Methontology propõe quatro relações taxonômicas, a saber: **subclasse de** (*subclass-of*), **decomposição disjunta** (*disjoint-decomposition*), **decomposição exaustiva** (*exhaustive-decomposition*) e **partição** (*partition*) ou **partitiva**.

A Figura 58 ilustra a taxonomia de conceitos sobre vôos, na qual aparecem relações de *subclasse de* como Vôo da *American Airlines*, Vôo da Ibéria e Vôo da *British Airways* sendo subclasses da classe Vôo. Isso implica em dizer que toda instância de Vôo da Ibéria, por exemplo, é também uma instância de Vôo. A Figura também mostra uma relação de *decomposição disjunta* do conceito Vôo da *British Airways* em que o conjunto de suas subclasses não possui instâncias em comum. A decomposição disjunta acontece nesse caso porque nenhum vôo pode ser simultaneamente um vôo BA0068, um vôo BA0066 e um vôo BA0069. Além disso, podem existir instâncias do conceito Vôo da *British Airways* que não sejam instâncias das três subclasses.

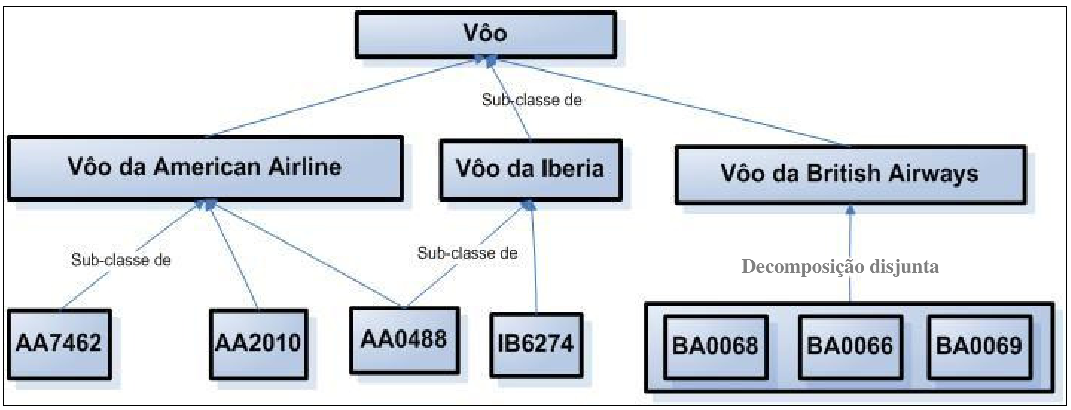


Figura 58 – Taxonomia de conceitos de vôos
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.134)

A Figura 59 ilustra a relação de decomposição exaustiva entre a classe Pacote de viagem e as subclasses Viagem econômica, Viagem executiva e Viagem de luxo. Tal relação implica em dizer que o conjunto de subclasses **cobre** o conceito de hierarquia superior, ou, em outras palavras, não existe instância na classe de hierarquia superior que não seja instância em pelo menos um dos conceitos na decomposição. Por exemplo, um pacote de viagem deve pertencer a uma das categorias de viagem (econômica, executiva ou luxo). Além disso, pode acontecer de conceitos possuírem instâncias em comum como, por exemplo, uma viagem executiva poderia ser econômica ou muito cara.

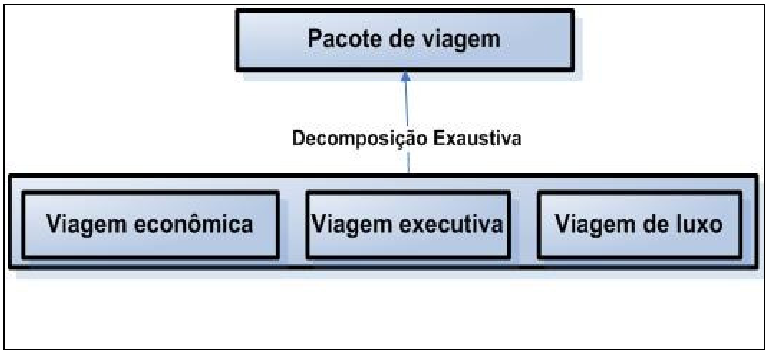


Figura 59 – Decomposição exaustiva na ontologia de viagem
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.135)

A Figura 60 ilustra a relação de partição ou partitiva envolvendo o conceito Vôo e suas subclasses Vôo Internacional e Vôo Doméstico. O conjunto de subclasses não compartilha instâncias em comum, mas cobrem o conceito de nível hierárquico superior, ou seja, não há instâncias em Vôo que não sejam instâncias de Vôo Internacional ou Vôo Doméstico.

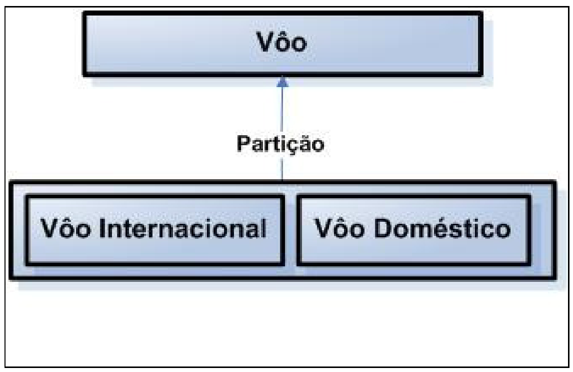


Figura 60 – Relação partitiva na ontologia de viagem
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.135)

Gomez-Perez, Fernandez e Vicente (1996) descrevem de forma detalhada os métodos aplicados nas atividades de elaboração das representações intermediárias relacionadas à descrição de conceitos. As representações intermediárias são importantes no sentido de especificar um conjunto de compromissos ontológicos estabelecidos entre os participantes do processo de construção de ontologias. Tais compromissos são fundamentais para a construção de uma conceitualização de senso comum que poderá ser compartilhada em um domínio de interesse.

As atividades envolvendo tais representações seriam: a) criar um *dicionário de dados*, o qual descreve e agrupa todos os conceitos úteis do domínio, seus significados, atributos, instâncias, etc; b) criar *árvores de classificação de conceitos*, conforme já foram elucidadas anteriormente; c) criar *tabelas de atributos de instância*, as quais fornecem informações sobre atributos e sobre seus valores nas instâncias; d) criar *tabelas de atributos de classe*, as quais descrevem os próprios conceitos, mas não suas instâncias; e) criar *tabelas de constantes*, as quais especificam informações relacionadas ao domínio de conhecimento que sempre têm o mesmo valor; f) criar *tabelas de*

fórmulas e *regras*, as quais explicitam o conhecimento do domínio por meio de regras e fórmulas; g) criar *árvores* ou *diagramas de classificação de atributos*, as quais apresentam graficamente a relação entre atributos e constantes na seqüência dedutiva de atributos de hierarquia superior, bem como a seqüência de fórmulas para tal dedução; e h) criar *tabelas de instâncias*, as quais definem as instâncias de conceitos.

A Tabela 16 (a, b e c) apresenta os métodos utilizados na condução das atividades envolvendo a construção das representações intermediárias e um conjunto de verificações em cada uma delas e entre elas, de forma a ajudar na detecção de inconsistências, redundâncias e omissão de conhecimento no domínio.

Representação Intermediária	Como construir	Verificação
Dicionário de dados	Para cada conceito identificado no domínio é necessário preencher os seguintes campos do dicionário: *nome do conceito*; *sinônimos* e *acrônimos* de nomes de conceitos; *descrição*, que fornece o significado de um conceito; *instâncias*, que incluem as instâncias de um conceito; *atributos de classe*, ou propriedades relevantes do conceito, que descrevem o próprio conceito; *atributos de instância*, ou propriedades relevantes que descrevem as instâncias de um conceito.	Verificar o conhecimento inserido dentro do dicionário buscando: a) garantir a integralidade do conhecimento sobre os conceitos envolvidos; b) determinar a granularidade dos conceitos cobertos pela ontologia; c) consistência sobre os atributos de classes e instâncias, em que devem fazer sentido para os conceitos envolvidos; d) assegurar a ausência de redundância e imprecisão sobre nomes e descrições de conceitos.
Árvore de classificação de conceitos	Após todos os conceitos terem sido direcionados ao dicionário de dados, o próximo passado seria ligar tais conceitos através das relações: *subclasse de*, *decomposição disjunta*, *decomposição exaustiva e partitiva*, conforme foram ilustradas nas Figuras 57, 58 e 59 anteriormente. O *ontologista* deve ser então capaz de classificar os conceitos em suas respectivas classes. Cada árvore concebida resulta em ontologias independentes e modularizadas no domínio de aplicação.	1) Verificar o conhecimento inserido dentro da árvore de classificação de conceitos. Tal representação possibilita visualizar graficamente a hierarquia de conceitos do dicionário de dados, tornando-se possível assegurar: a) que nenhum conceito seja repetido; b) ausência de ciclos entre conceitos; c) nenhuma árvore isolada a respeito dos conceitos relacionados. 2) Verificação cruzada entre representações intermediárias: todos os conceitos mostrados na árvore devem ser definidos no dicionário de dados. Caso contrário, o conceito deve ser adicionado no dicionário ou removido da árvore.

Tabela 16 (a) – Métodos para construção das representações intermediárias

Representação Intermediária	Como construir	Verificação
Tabelas de atributos de instância	1) Uma tabela deve ser criada para cada atributo de instância inserido no campo *atributos de instância* do dicionário de dados. 2) Para cada tabela incluir: *nome do atributo de instância*; *descrição* para fornecer o significado do atributo; *tipo de valor* (*float*, *boolean*, *string*) para determinar a restrição no momento do preenchimento do campo na classe; *unidade de medida* para valores numéricos (o símbolo "_" é usado para indicar campo sem valor e o símbolo "desconhecido" para campo desconhecido); *precisão* para valor numérico; *faixa de valores* para possíveis valores no atributo; *valores padrão*, caso sejam conhecidos; *cardinalidade* para especificar o número de valores do atributo; *atributos de instância inferidos*, *atributos de classe inferidos* e *constantes inferidas* (incluem o nome dos atributos de instância, de classe e constantes), que permitem inferir o valor do atributo em questão; *fórmula*, que faz referência cruzada com as tabelas de fórmulas, permitindo calcular o valor numérico do atributo; *a inferir*, que contém o nome dos atributos de instância cujos valores podem ser inferidos a partir do atributo em questão; e *referência* para registrar a fonte de informação (livro, especialista, etc).	1) Verificar o conhecimento dentro da tabela de atributos de instância de forma a garantir a consistência do conhecimento de senso comum para a ontologia construída. A saber: a) verificar a correspondência entre valores preenchidos nas tabelas de atributos de instância. Por exemplo: em uma tabela x, quando o campo *atributos de instância inferidos* for preenchido, as tabelas ou a tabela correspondente ao atributo indicado neste último campo deve possuir em seu campo *a inferir* o nome do atributo representado na tabela x; b) valores padrão devem ser preenchidos de acordo com o tipo de dado definido para o atributo. 2) Verificar o conhecimento entre representações intermediárias: a) verificar a correspondência entre as tabelas de atributos de instância criadas e os *atributos de instância* listados no dicionário de dados; b) verificar a correspondência de valores preenchidos nos campos *atributos de classe inferidos* e *constantes inferidas* da tabela de atributos de instância com o valor preenchido no campo *a inferir* das tabelas de atributos de classe e constantes.
Tabelas de atributos de classe	1) Uma tabela deve ser criada para cada conceito inserido no campo *nome do conceito* do dicionário de dados. 2) Para cada tabela especificar os seguintes campos: *nome do atributo de classe*; *nome do atributo relação*, que identifica o atributo participante de um relacionamento; *relacionamento lógico*, que uni dois conceitos através de operadores lógicos; *valor* do atributo de classe; *unidade de medida* para valores numéricos; *a inferir*, que contém o nome dos atributos de instância cujos valores podem ser inferidos a partir do valor do atributo de classe em questão; *referências* para registrar a fonte de conhecimento.	1) Verificar o conhecimento dentro da tabela de atributos de classe de forma a evitar omissões. 2) Verificar o conhecimento entre representações intermediárias: a) para todos os conceitos definidos em *atributos de classe* no dicionário de dados, deve existir uma tabela de atributos de classe, e para cada tabela de atributos de classe, deve existir um conceito no dicionário; b) para todos os atributos de classe especificados no campo *atributos de classe* no dicionário, deve existir uma tabela de atributos de classe correspondente, e vice versa.

Tabela 16 (b) – Métodos para construção das representações intermediárias

Representação Intermediária	Como construir	Verificação
Tabela de constantes	Para cada árvore de classificação de conceito, o *ontologista* deve identificar um conjunto de constante e descrevê-las a partir das seguintes propriedades: *nome da constante*; *descrição*, que indica o significado da constante; *valor da constante*; *unidade de medida* para valores numéricos; *a inferir* com a relação dos atributos de instância cujos valores podem ser inferidos a partir do valor da constante em questão; *referências* das fontes de conhecimento, nas quais as constantes foram identificadas.	1) Verificar o conhecimento dentro da tabela de constantes: se uma constante aparecer nas tabelas de atributos ou de fórmulas, deve também ser inserida na tabela de constantes. 2) Verificar o conhecimento entre representações intermediárias: conforme especificado na representação de tabelas de atributos de instância, constantes são usadas para inferir valores de atributos de instâncias. Nesse sentido, as verificações feitas em tal representação já seriam suficientes e válidas.
Tabelas de fórmulas	A definição padrão para fórmulas inclui: *nome da fórmula*; *atributo inferido*, que identifica o atributo de instância que é calculado a partir da fórmula; *fórmula*, que inclui a fórmula matemática para ser usada no cálculo do atributo inferido; *descrição* para incluir os fundamentos teóricos da fórmula; *atributos de instância básicos*, que inclui uma lista de atributos de instância usados na fórmula para calcular o atributo inferido; *atributos de classe básicos*, que inclui uma lista de atributos de classe usados na fórmula para calcular o atributo inferido; *constantes*, que inclui uma lista de constantes usadas na fórmula para calcular o atributo inferido; *precisão* com que o número deve ser calculado (ex.: dois números decimais); *restrições* para impor algumas condições às fórmulas; *referências* de fontes de conhecimento utilizadas na elucidação das fórmulas.	1) Verificar o conhecimento dentro da tabela de fórmulas: a) todos os atributos que aparecem no lado direito de uma fórmula devem aparecer nos campos *atributos de instância básicos*, *atributos de classe básicos* e *constantes*; b) se a definição especifica alguma restrição, outras fórmulas deveriam cobrir situações alternativas. 2) Verificar o conhecimento entre representações intermediárias: a) verificar se o campo *fórmula* especificado nas tabelas de atributos de instância corresponde à tabela de fórmula referenciada; b) verificar na tabela de fórmulas, se o campo *atributo inferido* encontra-se preenchido com o nome de um atributo de instância existente na tabela correspondente; c) verificar se o conteúdo dos campos *atributos de instância inferidos*, *atributos de classe inferidos* e *constantes inferidas* na tabela de atributos de instância corresponde com o conteúdo do campo *fórmula* da tabela de fórmulas.

Tabela 16 (c) – Métodos para construção das representações intermediárias

Representação Intermediária	Como construir	Verificação
Árvore de classificação de atributos	Após o *ontologista* verificar a consistência e a integralidade das tabelas de atributos de instância, atributos de classe, constantes e fórmulas, o próximo passo é desenvolver as árvores de classificação de atributos. Usar os campos *atributos de instância inferidos*, *atributos de classe inferidos* e *constantes inferidas*, *a inferir* e *fórmula* de cada tabela de atributos de instância. Tais campos serão a base para desenvolver as árvores de classificação de atributos.	1) Verificar o conhecimento dentro da árvore de classificação de atributos: a) se alguns atributos e constantes na seqüência dedutiva de atributos de hierarquia superior estão ausentes, errados ou não estão sendo utilizados para inferir valores de outros atributos; b) se algumas fórmulas estão ausentes ou sendo usadas erroneamente; c) se não há ciclos na árvore, ou seja, um atributo de instância não é capaz de inferir ele próprio. 2) Verificar o conhecimento entre representações intermediárias: a) qualquer modificação (inserção, remoção) nas árvores deve-se modificar também as tabelas envolvidas, isto é, tabelas de atributos, constantes e fórmulas; b) se qualquer atributo for incluído na árvore e não estiver presente no dicionário de dados, faz-se necessário incluí-lo no mesmo e criar a sua tabela correspondente.
Tabelas de instâncias	1) Criar uma tabela de instância para cada instância identificada no dicionário. 2) Os seguintes campos compõem a tabela: *nome da instância*; *descrição*, que fornece o significado da instância; *atributos* para preenchimento na instância em função das classes, nas quais as instâncias estão ligadas; *valores* desses atributos, que podem ser numérico, conjunto de valores, “desconhecido”.	1) Verificar o conhecimento dentro da tabela de instância: a) consistência entre atributos e seus valores; b) ler a descrição da instância de maneira cuidadosa para evitar redundâncias entre instâncias. 2) Verificar o conhecimento entre representações intermediárias: a) para todas as instâncias identificadas no campo *instância* do dicionário de dados existe uma tabela de instância correspondente e vice-versa; b) manter consistência entre os valores de atributo da instância e o campo *tipo de valor* definido em sua tabela de atributos de instância; c) manter consistência entre o número de valores declarado no atributo e o campo *cardinalidade* definido em sua tabela de atributos de instância; d) propagação de atributos usando herança de classes, ou seja, a instância seria preenchida com valores concretos de acordo com a árvore de classificação de conceitos.

Tabela 16 – Métodos para construção das representações intermediárias
Fonte: Adaptado de Gomez-Perez, Fernandez e Vicente (1996)

Vicente (1997)[126] apud Fernandez, Gomez-Perez e Juristo (1997) apresenta procedimentos para tratar os *verbos*, os quais são as ações no domínio e são descritos através das seguintes representações: a) dicionários de verbos: descrevem o significado dos verbos; b) tabelas de condições: descrevem um conjunto de condições a serem satisfeitas antes da execução da ação. Pode-se notar que tanto a árvore de classificação de conceitos como o diagrama de verbos utiliza as tabelas de fórmulas e de regras.

A relação completa dos documentos resultantes da execução das atividades de descrição de conceitos, expostas na Tabela 16 (a, b e c), pode ser encontrada em Gomez-Perez, Fernandez e Vicente (1996). E exemplos de atividades formalizadas (ou representações intermediárias) a partir da experiência dos autores na construção de ontologias podem ser encontrados em Fernandez, Gomez-Perez e Corcho (2004, pgs. 130-141).

4.5.4 Formalização

Gomez-Perez, Fernandez e Vicente (1996) indicam formalizar o modelo conceitual em uma linguagem formal como ontolingua e lógica descritiva. Entretanto, Fernandez, Gomez-Perez e Corcho (2004) assinalam que a formalização não é procedimento obrigatório na Methontology, pois ferramentas como a ODE, por exemplo, já geram o código através da exportação da especificação da ontologia na linguagem de representação utilizada na ferramenta.

4.5.5 Implementação

A fase de implementação requer o uso de um ambiente que suporte as meta-ontologias e ontologias selecionadas na fase de integração. O resultado desta fase é a codificação da ontologia em uma linguagem formal como Prolog, Ontolingua, dentre outras.

O modelo conceitual ou a conceitualização pode ser automaticamente implementado em várias linguagens usando tradutores (FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004). Alguns critérios para a escolha de ambientes para desenvolvimento de ontologias são enumerados como se segue: i) analisador léxico e sintático para garantir correções de possíveis erros; ii) tradutores para garantir a portabilidade de definições em outras linguagens; iii) um editor para adicionar, remover ou modificar definições; iv) um aplicativo de pesquisa para examinar bibliotecas de ontologias e suas definições; v) uma máquina de busca para examinar as definições mais apropriadas; vi) avaliadores para detectar conhecimento incompleto, inconsistente e redundante; vii) gerenciador automático para verificar a inclusão, remoção ou modificação de definições já existentes.

[126] VICENTE, A. *Concepualizacion de verbos en ontologias de dominio*. Tésis de Master em Ingeniería del Conocimiento. Facultad de Informatica. Universidad Politécnica de Madrid. 1997.

4.5.6 Integração

Esta fase considera a reutilização de conceitos já existentes em outras ontologias. A proposta é examinar meta-ontologias (como, por exemplo, a Cyc) e selecionar aquelas que melhor se ajustarem a conceitualização.

O resultado é um documento de integração que descreve, para cada termo aproveitado, o nome do termo e a definição do termo na meta-ontologia, o nome da meta-ontologia e o nome do termo na ontologia em construção.

4.5.7 Avaliação

Esta fase consiste em executar um julgamento técnico das ontologias, de cada fase e entre fases do ciclo de vida da ontologia. A atividade de avaliação divide-se em *verificação* e *validação*.

A tarefa de *verificação* consiste em verificar tecnicamente a consistência da ontologia, do ambiente de *software* associado e da documentação ao longo do ciclo de vida da ontologia. A tarefa de *validação* consiste em garantir que a ontologia, o *software* e a documentação correspondam ao sistema planejado. Propostas de avaliação têm sido desenvolvidas (GÓMEZ-PEREZ, 1994; GÓMEZ-PEREZ, 1999) por pesquisadores que colaboraram na especificação da Methontology. Tais propostas contribuem no amadurecimento da metodologia no sentido de assegurar que ontologias construídas a partir da Methontology consigam atender aos propósitos determinados.

4.5.8 Documentação

Segundo os autores da metodologia, não existe um guia consensual de como documentar a construção de ontologias. Em muitos casos, a documentação disponível contempla apenas o código da ontologia e um texto em linguagem natural anexado às definições formais (FERNANDEZ, GOMEZ-PEREZ E JURISTO, 1997). Gomez-Perez, Fernandez e Vicente (1996) destacam a importância da geração de documentação na fase de conceitualização. Segundo os autores, a documentação nessa fase possibilita: a) examinar se a ontologia é útil e utilizável para um domínio sem precisar tomar conhecimento de seu código fonte; e b) comparar o escopo da ontologia com a sua integralidade, reusabilidade e compartilhamento através da análise do conhecimento expressado em cada representação intermediária.

A Methontology pretende incluir a documentação como uma atividade a ser realizada durante todo o processo de desenvolvimento da ontologia, ou seja, para cada fase concluída deve ser desenvolvida uma documentação registrando todo o conhecimento pertinente ao que foi feito.

Tabela sinóptica da metodologia Methontology

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Planejamento das atividades; **controle** das tarefas e garantia da **qualidade**.	Ausente	Definição do **propósito**, do **grau** de **formalização** e do **escopo**.	Concepção da **conceitualização** através das **representações intermediárias**: i) glossário preliminar de termos; ii) árvores de classificação de conceitos; iii) diagramas de verbos; iv) dicionário de dados; v) tabelas de atributos de instâncias; vi) tabelas de atributos de classe; vii) tabela de constantes; viii) tabelas de fórmulas e regras; ix) árvores de classificação de atributos; e x) tabelas de instância.	Formalizar o modelo conceitual em uma **linguagem formal** como ontolingua e lógica descritiva. Entretanto, a formalização não é procedimento obrigatório, pois ferramentas como a ODE, por exemplo, já geram o código através da exportação da especificação da ontologia na linguagem de representação disponível.	**Ambientes** de desenvolvimento **adequados** à meta-ontologias e ontologias selecionadas na fase de integração. Uso de linguagens como Prolog, Ontolingua, dentre outras.	Orientação quanto a **ajustes necessários** nas **representações intermediárias** (ver tabela 16)	O resultado é um **documento de integração** que descreve, para cada termo aproveitado, o nome do termo e a definição do termo na meta-ontologia, o nome da meta-ontologia e o nome do termo na ontologia em construção.	Divide-se em **verificação** e **validação** da ontologia.	Recomendada em **cada fase** do ciclo de vida.

Tabela 17 – Tabela sinóptica da metodologia Methontology

4.6 Método Sensus

A ontologia SENSUS foi desenvolvida pelo grupo de linguagem natural *Information Sciences Institute* - ISI com o propósito de ser usada para fins de processamento de linguagem natural. A ontologia SENSUS possui aproximadamente 70.000 conceitos organizados em uma hierarquia, de acordo com seu nível de abstração que vai de médio a alto. No entanto, sua estrutura não contempla termos específicos de um domínio (SWARTOUT et al., 1996). Para tal, os termos de domínios específicos são ligados à ampla ontologia SENSUS, de forma a construir ontologias para domínios particulares, como é o caso do domínio para planejamento de operação militar aérea (VALENTE et al, 1999).

O método Sensus, baseado na ontologia SENSUS, propõe alguns processos para estabelecer as ligações entre os termos específicos e os termos da ontologia de alto nível (SWARTOUT et al., 1996). O resultado de tal processo é uma estrutura de uma nova ontologia, que é generalizada automaticamente através de uma ferramenta denominada OntoSaurus (SWARTOUT et al., 1996; FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004), a qual será descrita mais adiante nessa seção.

De acordo com o método, os processos envolvidos na construção da ontologia de um domínio específico seriam: i) identificar termos-chave do domínio; ii) ligar manualmente os termos-chave à ontologia SENSUS; iii) adicionar caminhos até o conceito de hierarquia superior da Sensus; iv) adicionar novos termos para o domínio; e v) adicionar subárvores completas. Swartout et al. (1996) assinalam que a principal vantagem do método Sensus é que duas ontologias poderiam ser construídas de forma independente e posteriormente torná-se-ía possível integrar as suas terminologias a partir da ontologia SENSUS, a qual possui uma ampla cobertura de termos organizados. Desse modo, a abordagem Sensus promove compartilhamento de modelos.

As seções a seguir apresentam os processos de construção, propostos pelo método Sensus, mapeados nas seguintes categorias de análise: *modelagem conceitual*, *integração*, *formalização* e *implementação*.

4.6.1 Modelagem Conceitual e Integração

O conjunto de cinco passos proposto pelo método Sensus contempla tanto a fase de modelagem conceitual de um domínio específico quanto à fase que diz respeito à integração entre conceitos de uma ou várias ontologias com a SENSUS ou entre elas. Dessa maneira, tais fases são apresentadas em conjunto nessa seção, tendo em vista que a atividade de integração faz parte dos processos integrais, ou seja, pode ser realizada em qualquer fase do ciclo de desenvolvimento da ontologia. A condução dos passos é feita de maneira semi-automática, ou seja, alternando-se entre

atividades manuais e com uso de ferramenta automatizada como apoio à concepção do modelo, conforme será visto adiante.

O primeiro passo para o desenvolvimento do modelo sobre um domínio específico seria identificar os termos na SENSUS que são relevantes para tal domínio, além de identificar os termos relevantes (também chamados de "semente") para o domínio (veja exemplos de termos relevantes na Figura 61). Em tal atividade, os autores do método (SWARTOUT et al., 1996) vêm usando como guia a lingüística e técnicas de agrupamento (do inglês, *clustering*) de conceitos.

O segundo passo objetiva ligar manualmente os termos do domínio específico com a SENSUS. A Figura 61 ilustra dois conceitos de domínio, Vôo Europa-África e Vôo Europa-América, sendo ligados à ontologia SENSUS como hipônimos[127] (subclasses) de Vôo Internacional; e outros dois conceitos, Vôo Londres-Liverpool e Vôo Madri-Barcelona, ligados como hipônimos a Vôo Doméstico.

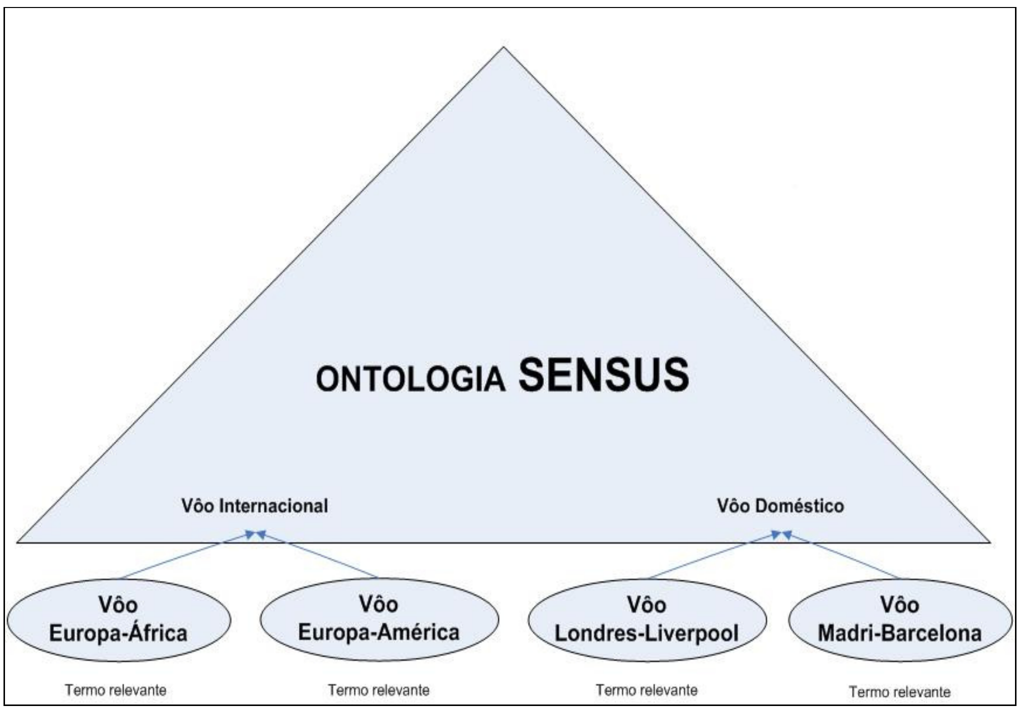


Figura 61 – Ligação de termos relevantes para o domínio à ontologia SENSUS
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.144)

O terceiro passo compreenderia incluir todos os conceitos que estiverem no caminho compreendido entre os termos relevantes para o domínio e o conceito de hierarquia superior (também chamado de "raiz") da SENSUS. Tal passo demandaria o uso de uma ferramenta, apresentada na Figura 62, para visualização dos termos relacionados na árvore de conceitos da

[127] Diz-se de vocábulo ou sintagma de sentido mais específico em relação ao de um outro mais geral, em cuja classe está contido. Por exemplo, poltrona é hipônimo de assento; leão é hipônimo de animal.

ontologia. A Figura 63 ilustra o conjunto de conceitos compreendido entre o alto nível e o nível específico da hierarquia.

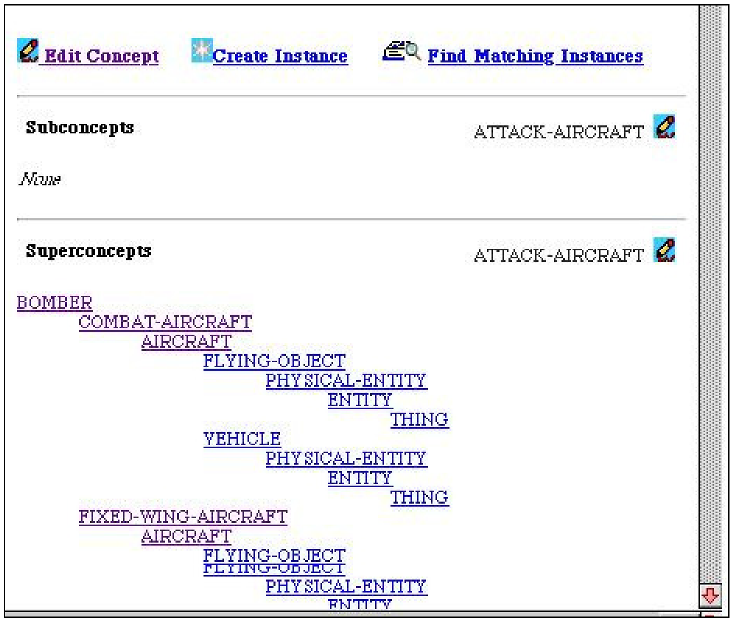


Figura 62 – Visualização de termos no OntoSaurus
Fonte: Swartout et al., (1996)

O quarto passo demanda a adição de outros termos relevantes, ilustrados na Figura 63, no domínio específico que não foram considerados no passo dois, isto é, na ligação manual com a ontologia Sensus. Desta forma, os passos dois e três precisariam ser repetidos de forma a organizar novamente a hierarquia de conceitos.

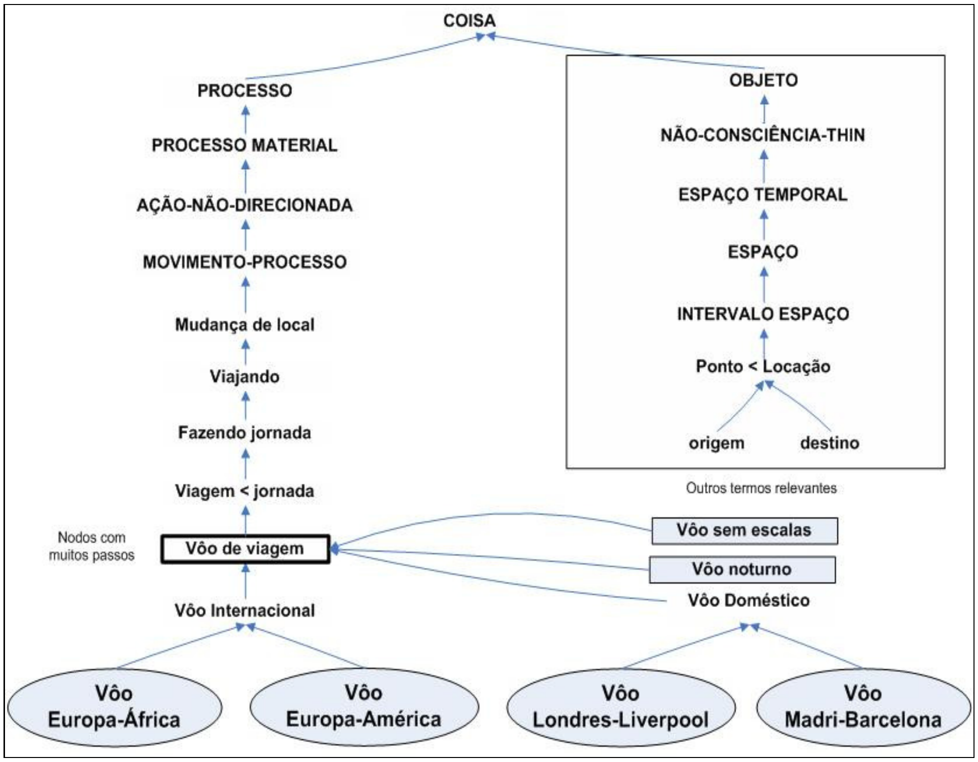


Figura 63 – Apresentação do resultado do passo 3, 4 e 5
Fonte: Adaptado de Fernandez, Gomez-Perez e Corcho (2004, p.145)

E, finalmente, o quinto passo, que propõe adicionar, também de maneira manual, *subárvores* completas sob os nodos que têm muitos caminhos ou passos, como no caso do exemplo apresentado na Figura 63, em que uma *subárvore* completa é adicionada sob o conceito Vôo de viagem.

4.6.2 Formalização e Implementação

Para a concepção de uma ontologia de domínio, o método Sensus utiliza a ferramenta OntoSaurus (um servidor web de ontologias), a qual foi concebida para fins de desenvolvimento, gerenciamento e manutenção de bases de conhecimento, através da linguagem de programação LOOM (SWARTOUT et al., 1996). A ferramenta possui as seguintes características, que viabilizam a construção de ontologias: i) fornece acesso para ontologias; ii) permite apresentação da hierarquia da ontologia em formato HTML; iii) permite a edição de ontologias *on-line* para **formalização** de conceitos (editor apresentado na Figura 64), incluindo definições conceituais através de *hiperlinks* para outros conceitos e relações, isto é, através de uma rede semântica conceitual; iv) definição de regras (através de *slots*) e instâncias de conceitos na base de conhecimento; v) fornece vários tradutores (para Ontolingua, KIF e C++), que podem viabilizar a interoperabilidade entre sistemas heterogêneos que fazem uso de ontologias; e iv) pelo fato de incorporar a linguagem LOOM torna-se possível usufruir da capacidade de inferência de tal linguagem.

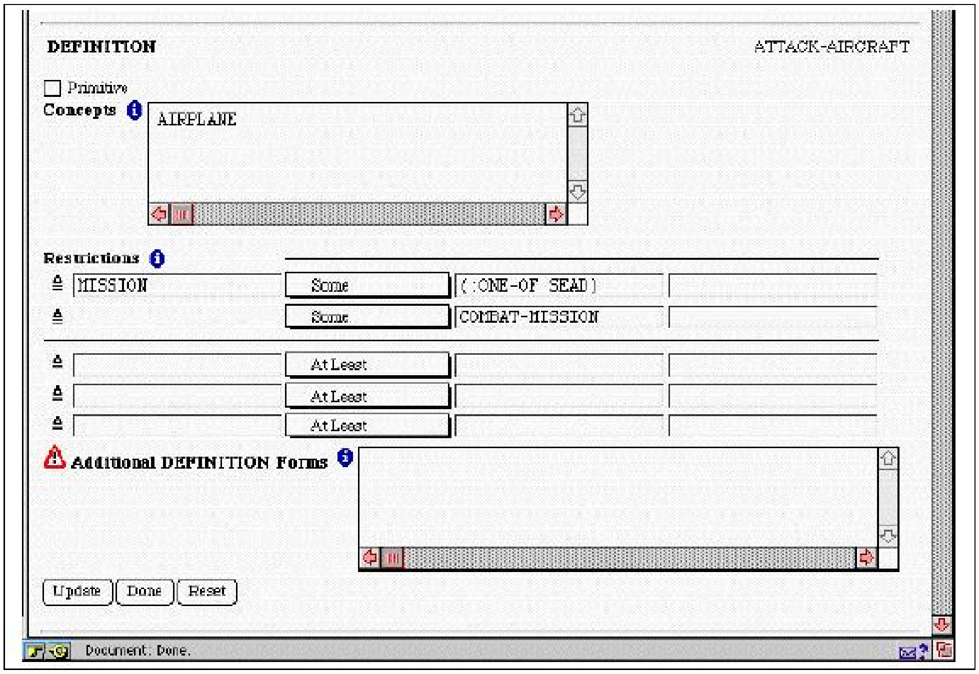


Figura 64 – Formulário de edição de conceitos no OntoSaurus
Fonte: Swartout et al., (1996)

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Ausente	Ausente	i) **identificar termos-chave** do domínio; ii) **ligar manualmente** os termos-chave à ontologia SENSUS; iii) **adicionar caminhos** até o conceito de hierarquia superior da Sensus; iv) **adicionar novos termos** para o domínio; e v) **adicionar subárvores** completas.	**Rede semântica**	Utiliza a ferramenta **OntoSaurus**, a qual foi concebida para fins de desenvolvimento, gerenciamento e manutenção de bases de conhecimento, através da **linguagem** de programação **LOOM**.	Ausente	**Ligar** manualmente os **termos-chave** à **ontologia SENSUS**.	Ausente	Ausente

Tabela 18 – Tabela sinóptica do método Sensus

4.7 Método 101

O método 101 foi concebido por Natalya F. Noy e Deborah L. McGuinness a partir da experiência no desenvolvimento da ontologia de vinhos e alimentos, utilizando o ambiente de edição de ontologias *Protégé-2000*. Noy e McGuinness (2001, p.2) afirmam que "não existe uma única metodologia correta para construção de ontologias "[128] e que não pretendem definir tal metodologia. A idéia do guia é apresentar a experiência das autoras no desenvolvimento de uma ontologia de modo a ser útil para o desenvolvimento de outros projetos.

Segundo as autoras do método, algumas idéias para construção de ontologias presentes no guia (NOY e McGUINNESS, 2001) tiveram origem na literatura do paradigma orientado a objetos (POO), contudo ressaltam que no desenvolvimento de ontologias, o projeto de classes e relações difere-se da programação orientada a objetos nos seguintes aspectos: a) reflete a estrutura do mundo real por meio de conceitos e não a estrutura de dados e código como no POO; e b) não se preocupa com a representação física das coisas e sim com a semântica de tais coisas, divergindo do POO, que descreve a representação física dos dados ou objetos.

O método 101 propõe basicamente quatro atividades para o desenvolvimento de uma ontologia: i) definir classes na ontologia; ii) organizar as classes em uma taxonomia; iii) definir *slots* e descrever seus valores permitidos (facetas[129]); e iv) adicionar valores de *slots* para as instâncias. Tais atividades implicam em decisões de modelagem, dentre as quais o método busca enfatizar (conforme será visto mais adiante), além de se encontrarem dentro de um processo iterativo de um ciclo de vida de ontologia. Para tal, sete passos, apresentados na Figura 65, são considerados importantes no processo de construção de ontologias, de acordo com o método.

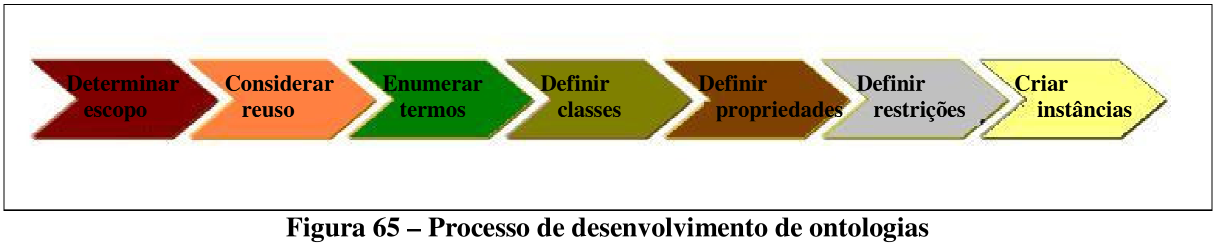


Figura 65 – Processo de desenvolvimento de ontologias
Fonte: Adaptado de Noy e McGuinness (2001)

Diante às atividades enumeradas anteriormente, as seções a seguir objetivam apresentar em maior detalhe os passos propostos no método 101 a partir do mapeamento das seguintes categorias de análise: *especificação de requisitos*, *modelagem conceitual*, *formalização*, *implementação*, *integração* e *documentação*.

[128] "[…] there is no single correct ontology-design methodology […]"

[129] Os conceitos de *slot* e faceta foram elucidados na fundamentação teórica desta pesquisa, especificamente na seção 2.3.1.2 sobre *frames*, tendo-se em vista que o *Protégé-2000* é uma ferramenta baseada em *frame*.

4.7.1 Especificação de Requisitos

A fase de especificação de requisitos é o ponto de partida para o desenvolvimento de uma ontologia, segundo o método 101. A primeira atividade é determinar o domínio e o escopo da ontologia, sintetizada nas seguintes questões:

- Que domínio a ontologia vai cobrir?
- Para que nós vamos utilizar a ontologia?
- Que tipo de questões as informações na ontologia vão responder?
- Quem irá usar e manter a ontologia?

As respostas a estas questões podem mudar ao longo do tempo, mas podem ajudar a limitar o escopo do modelo. No intuito de determinar o escopo da ontologia, questões de competência (GRUNINGER e FOX, 1995) são formuladas de forma a especificar o que a base de conhecimento, baseada na ontologia, deverá ser capaz de responder. As questões de competência para o domínio de vinhos e alimentos são explicitadas abaixo:

- Quais características do vinho eu devo considerar na escolha?
- *Bordeaux* é um vinho tinto ou branco?
- *Cabernet Sauvignon* combina com pescado?
- Qual é a melhor escolha de vinho para carnes grelhadas?
- Quais características de um vinho afetam sua combinação com um prato?
- O sabor ou a cor de um vinho específico muda de acordo com a safra?
- Quais foram as melhores safras para *Napa Zinfandel*?

Considerando tais questões, a ontologia deverá incluir informações relativas a características de vários vinhos e tipos de vinhos, anos de safras e classificações de alimentos que combinam com um tipo específico de vinho.

4.7.2 Modelagem Conceitual e Formalização

No método 101, a fase de modelagem conceitual e a fase de formalização são executadas simultaneamente. O método foi concebido a partir da experiência de construção da ontologia de vinhos e alimentos na ferramenta *Protégé-2000*, e as atividades envolvendo modelagem e formalização do modelo são executadas diretamente em tal ferramenta. Portanto, o conhecimento da

ontologia é formalizado em uma linguagem de representação baseada em *frames*, conforme pode ser visualizado nas Figuras[130] 67 e 68. As seções a seguir apresentam os passos que possibilitam a construção da conceitualização da ontologia: i) enumerar os termos importantes; ii) definir classes e hierarquia das classes; iii) definir as propriedades das classes; iv) definir as facetas dos *slots*; e v) criar instâncias. O guia que descreve o método (NOY e McGUINNESS, 2001) fornece uma série de recomendações para a fase de modelagem conceitual no momento de definir as classes, a hierarquia das mesmas, as propriedades e a terminologia para os conceitos. Tais recomendações são descritas nas próximas seções.

4.7.2.1 Enumerar os termos importantes

Após a especificação de requisitos ter sido definida através do escopo ou questões de competência da ontologia, o passo seguinte é **enumerar os termos importantes** para a ontologia, utilizando-se das seguintes questões: a) quais são os termos que nós estamos tratando? b) quais são as suas propriedades? c) o que nós queremos dizer sobre estes termos? Exemplos de termos relacionados ao domínio de vinhos e alimentos podem ser listados: *wine*, *grape*, *winery*, *location*, *wine color*, *wine body*, *wine flavor*, *sugar content*, *white wine*, *red wine*, *fish*, *read meat*.

4.7.2.2 Definir classes e hierarquia das classes

Após a identificação dos termos e suas propriedades, os próximos dois passos são considerados os mais importantes no processo de construção da ontologia: a **definição de classes** e a **hierarquia de classes**.

Para a definição das classes da ontologia, os termos identificados anteriormente são selecionados para descrever os objetos que envolvem o domínio. Tais objetos são agrupados de acordo com propriedades similares de forma a conceber as classes. Estas, por sua vez, podem ter instâncias como, por exemplo, uma garrafa de vinho e são organizadas em uma hierarquia taxonômica.

O guia assinala que não há uma única hierarquia de classe correta para qualquer domínio, mas sim melhores práticas. A hierarquia depende dos possíveis usos da ontologia, do nível de detalhe necessários à aplicação, das preferências pessoais e dos requisitos para compatibilidade com outros modelos. No método 101, para o desenvolvimento de uma hierarquia de classes as abordagens *top-down*, *bottom-up* e *middle-out*, vislumbradas em Uschold e Gruninger (1996), são consideradas. A Figura 65 mostra o processo de desenvolvimento de nível superior (*top-down*) cujo

[130] As figuras foram generalizadas através do *plugin Onto Viz* do *Protégé-2000* (NOY e GUINNESS, 2001).

início parte da classe mais geral *Wine* para então especializá-la através da criação das subclasses *White wine*, *Rosé wine* e *Red Wine* e assim sucessivamente. O desenvolvimento de nível inferior (*bottom-up*) inicia o processo com a definição de classes mais específicas como, por exemplo, a definição das classes *Pauillac* e *Margaux*, as quais são agrupadas na superclasse *Medoc*, que por sua vez é subclasse de *Red Bordeaux*. E, por fim, o desenvolvimento que combina as abordagens *top-down* e *bottom-up*. Os conceitos mais proeminentes são identificados e definidos para posteriormente serem generalizados e especializados de maneira apropriada. De acordo com a Figura 66, o processo de construção da taxonomia poderia iniciar com a definição de poucos conceitos gerais como *Wine* e com poucos conceitos específicos como *Margaux*, para então relacioná-los com conceitos de nível central (*middle-out*) como *Medoc*.

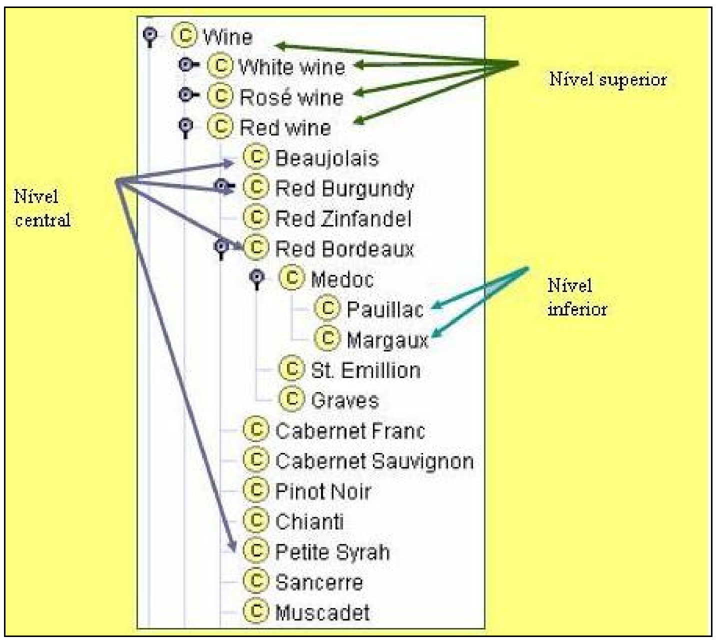


Figura 66 – Os níveis da taxonomia de vinhos
Fonte: Adaptado de Noy e McGuinness (2001, p.7)

Para as autoras do método, a melhor abordagem a ser usada vai depender da visão pessoal da pessoa que está modelando o domínio. Geralmente, a abordagem *middle-out* é considerada a mais fácil pelos *ontologistas*, tendo em vista que conceitos mais centrais tendem a ser mais descritivos perante o domínio.

No sentido de assegurar que a hierarquia de classes esteja correta, o método traz um conjunto de princípios apresentado na Tabela 19 (a).

Hierarquia	Princípios
Relação "é-um" ou "tipo-de"	A hierarquia de classe representa uma relação "é-um", portanto uma classe A é uma subclasse de B se toda instância de B é também uma instância de A. Ou, uma subclasse representa um conceito que é um "tipo de" conceito que a superclasse representa.
Transitividade da hierarquia de classes	Um relacionamento "é-um" transitivo: se B é uma subclasse de A, e C é uma subclasse de B, então C é uma subclasse de A. Por exemplo, na definição da hierarquia tem-se a classe vinho, a subclasse vinho branco como uma subclasse de vinho e a classe *Chardonnay* como uma subclasse de vinho branco. O relacionamento de transitividade implica que a classe *Chardonnay* é também uma subclasse de vinho. Uma **subclasse** considerada **direta** é aquela mais próxima da subclasse da classe e a recomendação é a seguinte: na hierarquia, **não existe classes** entre uma classe e sua subclasse direta. No exemplo, *Chardonnay* é uma subclasse direta de vinho branco e não é uma subclasse direta de vinho.
Evolução da hierarquia de classes	Manutenções podem ser necessárias na hierarquia de classes a partir de mudanças ocorridas no domínio. Para tal, manter a consistência da estrutura torna-se fundamental no processo de manutenção, pois novas classes podem surgir demandando uma classificação correta das mesmas na hierarquia.
Classes e seus nomes	Classes representam conceitos no domínio. O nome da classe pode mudar, mas o nome (termo) vai continuar a se referir ao mesmo conceito. Recomendação para definição dos termos para os conceitos: i) os nomes devem ser todos no singular (na prática, uma classe é mais usada no singular) ou todos no plural; ii) nomes sinônimos para o mesmo conceito não representam classes diferentes. Apenas um termo deve estar associado a uma classe. Muitos sistemas permitem associar uma lista de sinônimos a uma classe. Caso isso não seja possível, os sinônimos podem ser listados na documentação da classe; iii) utilização de prefixo e sufixo nos nomes a fim de diferenciar as classes dos *slots* (como ***has-***winery ou maker***-of***); iv) evitar abreviações nos nomes dos conceitos; v) as subclasses deveriam incluir o nome de suas classes na definição de seus nomes, por exemplo: as subclasses da classe *vinho* seriam nomeadas de *vinho* tinto e *vinho* branco.
Evitando ciclos de classes	Deve-se evitar hierarquia cíclica de classes como no exemplo: a classe A tem uma subclasse B e ao mesmo tempo B é uma superclasse de A. Isso significa declarar que as classes A e B são equivalentes, ou seja, todas as instâncias de A são instâncias de B e todas as instâncias de B também são instâncias de A.
Classes disjuntas	Classes são disjuntas se elas não podem ter instâncias em comum, além de não poderem ter nenhuma subclasse em comum. Por exemplo: vinho tinto, vinho branco e vinho rosé são classes disjuntas.
Herança múltipla	Alguns sistemas de representação de conhecimento permitem herança múltipla na hierarquia de classes. Isso acontece quando uma classe pode ter mais de uma superclasse. Por exemplo: o vinho do porto é ao mesmo tempo um vinho tinto e um vinho de sobremesa. Sendo assim, uma subclasse herda *slots* e facetas de todas as suas superclasses.

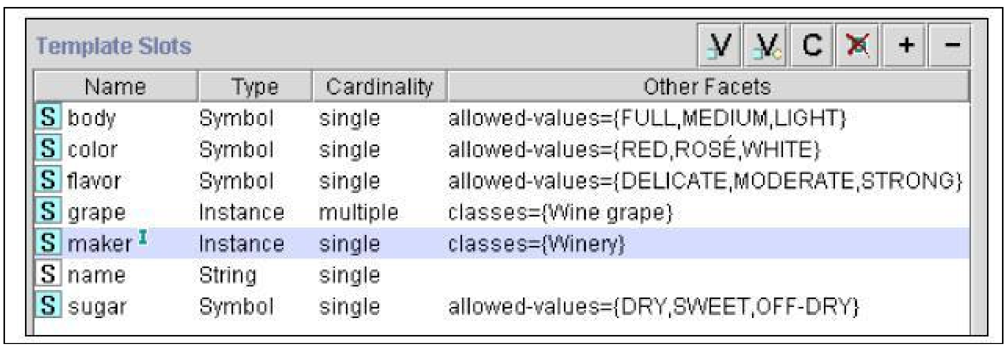
Tabela 19 (a) – Princípios para a definição de classes e hierarquia de classes

Hierarquia	Observação e Recomendação
Irmãos na hierarquia	Todos os irmãos na hierarquia devem estar no mesmo nível de generalidade. Por exemplo: vinho branco e *Chardonnay* não deveriam ser subclasses da mesma classe (no caso, vinho), visto que vinho branco é um conceito mais geral do que *Chardonnay*. Tal recomendação pode ser comparada, por exemplo, com as seções e subseções de um livro, que também se encontram no mesmo nível de generalidade.
Quantidade de subclasses para uma classe	Não existe uma regra geral para o número de subclasses diretas que uma classe deve ter. Entretanto, muitas ontologias bem estruturadas têm entre duas e doze subclasses diretas. Algumas recomendações: i) se uma classe tem apenas uma subclasse direta pode existir um problema de modelagem ou a ontologia não está completa. Por exemplo: se o único *Red Burgundy* que se tem é Côtes d'Or, porque criar uma hierarquia?; ii) se uma classe tem mais do que doze subclasses, então categorias intermediárias podem ser necessárias.
Quando inserir (ou não) uma nova subclasse	Recomendações: subclasses de uma classe usualmente i) têm propriedades adicionais que a superclasse não tem; ou ii) têm restrições diferentes das superclasses; ou iii) participam de diferentes relacionamentos do que as superclasses.
Uma nova classe ou um valor de propriedade	Recomendações a partir da questão: *de que maneira uma distinção de um conceito é considerada importante no domínio?* Se as distinções sobre o conceito não implicarem diretamente em suas relações com outros objetos, então não há necessidade de introduzir uma classe para tal conceito e sim criar uma propriedade para a classe envolvida com o mesmo. Em contrapartida, se uma distinção é importante no domínio e implicar em distinções para diferentes objetos e relações, então há necessidade de criar uma nova classe para a distinção. Exemplo: uma classe de vinho com as distinções branco, tinto e rosé. Tais distinções seriam consideradas propriedades da classe vinho ou seria necessário criar uma classe para cada uma delas?
Uma instância ou uma classe	A decisão depende do nível de granularidade determinada pela aplicação da ontologia. Instâncias individuais são conceitos mais específicos representados na base de conhecimento. Para tal identificação, recomenda-se analisar as questões de competência levantadas na fase de especificação de requisitos e identificar os conceitos mais específicos que constituirão respostas a tais questões. Tais conceitos podem ser bons candidatos para instâncias na base de conhecimento. Por outro lado, se conceitos possuem propriedades comuns e formam uma hierarquia natural, então se deve representá-los como classes que acomodarão instâncias individuais com tais propriedades.
Limitando o escopo	Uma ontologia não deve conter todas as informações possíveis sobre o domínio, ou seja, não é necessário especializar ou generalizar mais do que a aplicação requer; e não é necessário incluir todas as propriedades possíveis de uma classe (apenas as mais proeminentes para a aplicação). Por exemplo: a ontologia de vinho provavelmente não incluirá tamanho da garrafa, cor do rótulo e comida e vinho favorito.

Tabela 19 – Princípios para a definição de classes e hierarquia de classes
Fonte: Adaptado de Noy e McGuinness (2001)

4.7.2.3 Definir as propriedades das classes - *slots*

As classes por si só não oferecem respostas às questões de competência. Para tal, após a definição das classes deve-se **descrever a estrutura interna** das classes ou conceitos. As propriedades das classes podem ser extraídas da lista de termos (como por exemplo, *wine color*, *wine body*, *wine flavor*, *sugar content*), cujo passo foi descrito na seção 4.7.2.1. Para cada propriedade na lista será necessário determinar quais classes são descritas por tal propriedade. Em geral, existem vários tipos de propriedades de objetos que podem tornar-se *slots* na ontologia. *Slots* em uma definição de classe descreve atributos de instâncias da classe e as relações com outras instâncias como, por exemplo, um vinho tem cor, produtor, acidez, etc. Os tipos de propriedades podem ser: a) intrínseca e extrínseca: sabor (*flavor*) e nome (*name*) do vinho, respectivamente; b) partes: ingredientes em um prato; c) relações com outros objetos: produtora de vinho (*winery*). A Figura 67 mostra os *slots* da classe vinho.


Template Slots

Name	Type	Cardinality	Other Facets
S body	Symbol	single	allowed-values={FULL,MEDIUM,LIGHT}
S color	Symbol	single	allowed-values={RED,ROSÉ,WHITE}
S flavor	Symbol	single	allowed-values={DELICATE,MODERATE,STRONG}
S grape	Instance	multiple	classes={Wine grape}
S maker I	Instance	single	classes={Winery}
S name	String	single	
S sugar	Symbol	single	allowed-values={DRY,SWEET,OFF-DRY}

Figura 67 – Os *slots* da classe vinho e as facetas para tais *slots* construídos no *Protégé-2000*
Fonte: Noy e McGuinness (2001, p.8)

Uma subclasse herda todos os *slots* da superclasse, ou seja, se um vinho tem um nome e sabor, um vinho tinto também tem nome e sabor. Se uma classe tem múltiplas superclasses, ela herda *slots* de todas as superclasses.

Uma consideração relevante é o conceito de *slot* inverso, representado na Figura 67 pelo símbolo **I**. Um valor de um *slot* pode depender do valor de outro *slot*. Por exemplo, se um vinho foi produzido (*maker*) por uma vinícola, então a vinícola produz (*produces*) um vinho. As duas relações, *maker* e *produces*, são chamadas de relações inversas. Apesar do armazenamento de informação acerca dos *slots* inversos ser redundante, os mesmos permitem a apresentação de conhecimento nas duas direções, por exemplo, conhecimento sobre vinho de um lado, e conhecimento sobre vinícola de outro. Além do sistema de aquisição de conhecimento assegurar consistência a partir do preenchimento automático no valor da relação inversa.

4.7.2.4 Definir as facetas dos *slots*

As **restrições das propriedades**, também chamadas de facetas, descrevem ou limitam o conjunto de valores possíveis para um *slot*. As facetas mais comuns são *cardinalidade*, *tipo de valor* e *valor padrão*. Exemplos: uma vinícola tem exatamente uma localização; o nome de um vinho é do tipo *string*; o produtor de vinho (*wine producer*) é uma instância de Vinícola (*Winery*); o valor padrão para "*wine body*" poder ser "*full*".

A faceta *cardinalidade* define quantos valores um *slot* pode ter como nos seguintes casos:

- Cardinalidade mínima e máxima
  - N se o *slot* deve ter **ao menos** N valores ou **no máximo** N valores
  - 1 se o *slot* deve ter pelo menos um valor (requerido)
  - 0 se o *slot* é opcional
- Cardinalidade única e múltipla
  - Permite no máximo 1 valor (*single-valued slot*)
  - Permite qualquer número de valores (*multiple-valued slot*)

A faceta *tipo de valor* descreve quais tipos de valores podem preencher um *slot*, a saber:

- *String*: uma cadeia de caracteres
- *Number*: um inteiro ou ponto flutuante
- *Boolean*: tipo lógico verdadeiro/falso (true/false)
- *Enumerated*: lista de valores
- *Instance-type*: permite definir relações entre indivíduos. Os *slots* com esse tipo de faceta também devem definir uma lista de classes permitidas através das quais as instâncias podem ocorrer.

A classe (ou classes) em que os valores de *slot* podem pertencer é chamada de escopo (*range*) de um *slot*. A classe vinho é o *range* do *slot* produz. Já a classe cujo *slot* encontra-se ligado é chamada de domínio (*domain*) de um *slot* ou mais precisamente, a instância de classe (ou classes) que pode ter o *slot*. A classe vinícola é o *domain* do *slot* produz. Os conceitos *domain* e *range* foram fundamentados na seção 2.3.3.4.3 desta pesquisa, na qual vislumbrou o emprego de tais conceitos no intuito de restringir vocabulários concebidos na linguagem *Resource Description Framework Schema*.

A faceta *valores padrão* estabelece um valor assumido pelo *slot* na criação de uma instância. Tal valor pode ser mudado na faceta, diferentemente de um valor de *slot* pré-estabelecido em uma classe, cujo valor não pode ser mudado nas subclasses ou instâncias de tal classe.

4.7.2.5 Criar instâncias

E, finalmente, o último passo seria criar as instâncias das classes na hierarquia. Ao definir uma instância de uma classe deve-se:

- Escolher uma classe;
- Criar uma instância individual para a classe;
- Preencher valores no *slot*. Os valores precisam estar de acordo com as facetas.

No exemplo da Figura 68 é criada uma instância *Chateau Morgon Beaujolais* para representar um tipo específico de vinho *Beaujolais*. Tal instância pertence à classe *Beaujolais*, que por sua vez representa todos os vinhos *Beaujolais*. Os valores de *slot* (*body*, *color*, *flavor*, *tannin*, *grape*, *maker*, *area*, *sugar*) também são definidos para a instância em questão.



Figura 68 – Definindo instâncias para a classe Beaujolais no *Protégé-2000*
Fonte: Noy e McGuinness (2001, p.12)

4.7.3 Implementação

Ao finalizar todos os passos descritos anteriormente, o *ontologista* pode gerar automaticamente o código fonte da ontologia. O *Protégé* possui suporte a vários formatos de representação do conhecimento como XML *Schema*, RDF e OWL.

4.7.4 Integração

O método 101 considera a reutilização de ontologias com o propósito de: a) poupar esforço no desenvolvimento; b) interagir com as ferramentas que usam outras ontologias e c) utilizar ontologias validadas através do uso em aplicações. Entretanto, se um sistema de representação de

conhecimento não trabalhar diretamente com o formalismo em particular da ontologia em construção, faz-se necessário utilizar recursos de tradução de um formalismo para outro.

O método indica buscar ontologias em bibliotecas de reuso disponíveis na Web como DAML *ontology library* (www.daml.org/ontologies) e Ontolingua *ontology library* (www.ksl.stanford.edu/software/ontolingua/).

4.7.5 Documentação

Na definição das classes e dos *slots*, a ferramenta *Protégé* possibilita fazer a documentação descrevendo-os em linguagem natural, conforme pode ser visualizado na Figura 69. Além disso, as premissas relevantes sobre o domínio também devem ser registradas como, por exemplo, lista de sinônimos e definições sobre limitações de escopo. Segundo as autoras do método, a documentação é uma tarefa importante no processo de construção de ontologias, pois pode beneficiar os usuários no processo de comunicação acerca do projeto e também ser útil no processo de reutilização, visto que o conhecimento sobre os elementos da ontologia estaria documentado.

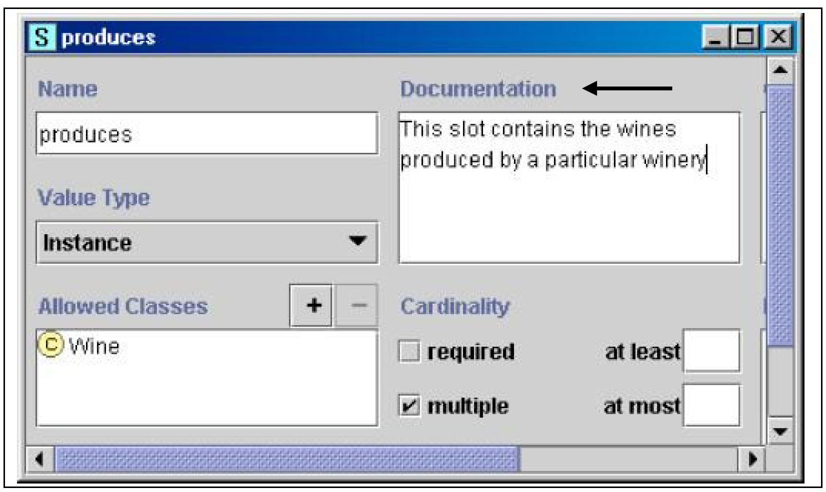


Figura 69 – Documentação feita no *slot* produces da ontologia de vinhos
Fonte: Noy e McGuinness (2001, p.10)

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Ausente	Ausente	Determinar o **domínio** e o **escopo** da ontologia. Definição de **questões de competência**	i) enumerar **termos**; ii) definir **classes**; iii) **organizar** as classes em uma **taxonomia**; iv) definir ***slots*** e **facetas**; v) adicionar **valores de *slots*** para as **instâncias**.	**Simultâneo** à fase de **modelagem conceitual** através de uma linguagem de representação baseada em *frames*.	Baseada no módulo de ***frames*** ou módulo **OWL** da ferramenta *Protégé*.	Ausente	Considera a **reutilização** de ontologias com o propósito de: a) poupar esforço no desenvolvimento; b) interagir com as ferramentas que usam outras ontologias e c) utilizar ontologias validadas através do uso em aplicações.	Ausente	Na **definição** das **classes** e dos ***slots***, a ferramenta *Protégé* possibilita fazer a documentação descrevendo-os em linguagem natural. Além disso, as **premissas relevantes** sobre o domínio também devem ser registradas como, por exemplo, lista de sinônimos e definições sobre limitações de escopo.

Tabela 20 – Tabela Sinóptica do método 101

4.8 Metodologia e Norma para construção de vocabulários controlados

A norma padrão proposta pela organização norte-americana *National Information Standards Organization* (ANSI, 2005) propõe as linhas gerais para construção, formatação e manutenção de vocabulários controlados monolíngües. Em relação às regras de construção, o padrão ANSI/NISO Z39.19-2005 permite construir diversos tipos de vocabulário controlado, incluindo tesauros, taxonomias, listas e anel de sinônimos, em uma ordem conhecida e estruturada com o intuito de disponibilizar claramente os relacionamentos de equivalência, associativos e hierárquicos, quando aplicáveis a cada tipo (ver apêndice B da norma, página 135). Essa flexibilidade é importante, pois permite uma melhor adaptação do instrumento às necessidades dos ambientes informacionais, como a Web, por exemplo.

Uma inovação da ANSI/NISO Z39.19-2005 em relação às normas anteriores é a inclusão da análise facetada, cuja base concentra-se no trabalho de Ranganathan e nos refinamentos feitos pelo *Classification Research Group* – CRG, na Inglaterra, conforme elucidado na fundamentação teórica desta pesquisa (seções 2.2.3.1 e 2.2.3.3.3). Na definição dos termos, a norma determina regras claras para o uso de qualificadores, além de possibilitar a inclusão de notas históricas para registrar todos os eventos ocorridos nos termos. Outra mudança importante é em relação às construções de apresentação de tesauros, que estão mais adaptados aos sistemas computacionais, inclusive a Web. Mas é importante assinalar que não são mencionados os formatos legíveis por máquinas como XML e RDF, por exemplo. Outra questão importante é a orientação sobre a interoperabilidade entre vocabulários controlados.

E, finalmente, o padrão ANSI/NISO Z39.19-2005 não propõe um ciclo de atividades para a construção de vocabulários controlados, conforme pode ser conferido em uma passagem da norma: “Este padrão deveria ser considerado como um conjunto de recomendações baseado em técnicas e procedimentos preferidos. Procedimentos opcionais são, entretanto, às vezes descritos, por exemplo, para a exibição de termos no vocabulário controlado”[131] (ANSI, 2005, p.1). A norma apenas recomenda uma série de tratamentos em relação ao termo frente à citação de atividades envolvidas no processo de construção, conforme poderá ser conferido nas subseções adiante.

Diante disso, para que fosse possível obter informações acerca dos procedimentos necessários à construção de vocabulários controlados, bem como das atividades envolvidas, foi utilizado como apoio o manual disponível no sítio da *Biblioteconomia, Informação e Tecnologia da Informação* – BITI (CAMPOS, GOMES e MOTTA, 2004) sobre elaboração de tesauros. Os procedimentos propostos no manual resumem-se em: i) *planejamento*; ii) *levantamento do*

[131] “This Standard should be regarded as a set of recommendations based on preferred techniques and procedures. Optional procedures are, however, sometimes described, e.g., for the display of terms in a controlled vocabulary”.

vocabulário; iii) *organização dos conceitos*; iv) *apresentação final*; e v) *critérios para avaliação*. Tal manual, mesmo focando em um tipo específico de vocabulário controlado, mostrou-se coerente em relação às recomendações tratadas na norma ANSI/NISO Z39.19-2005.

O propósito das seções seguintes é apresentar as atividades necessárias para se construir um vocabulário controlado frente à análise feita no manual da BITI e nas recomendações da norma ANSI/NISO Z39.19-2005. Buscando seguir a proposta de classificação de conteúdo por meio das categorias de análise, os objetos investigados foram mapeados utilizando-se as seguintes categorias: *gerenciamento do projeto*, *pré-desenvolvimento*, *especificação de requisitos*, *modelagem conceitual*, *formalização*, *manutenção*, *avaliação* e *documentação*. Ressalta-se que a categoria *formalização* foi aqui utilizada no contexto dos vocabulários controlados, segundo a ANSI/NISO Z39.19-2005.

4.8.1 Gerenciamento do projeto

O manual da BITI ressalta a necessidade de criação de um grupo específico para acompanhar, de forma permanente, o desenvolvimento do instrumento. Como atividades de planejamento de projeto, a metodologia adota: i) *levantamento das fontes*, incluindo literatura de divulgação científica, índices de publicações periódicas, índices de periódicos de resumo, revistas técnicas, teses e dissertações; ii) *decisão da forma de apresentação* para identificação dos requisitos para a seleção do *software* (no caso de se construir utilizando uma ferramenta como apoio); iii) *período de atualização* para averiguar a dinâmica de crescimento de cada área de assunto coberta pelo instrumento; iv) *seleção do software para construção* do instrumento, observando os requisitos de restrição à entrada de termos repetidos, criação de vários tipos de relacionamentos, produção de relatórios hierárquicos e alfabéticos e correção e alteração em cascata; e v) *constituir um grupo específico* para atuar na manutenção do instrumento em função de possíveis atualizações.

Para a norma ANSI/NISO Z39.19-2005, a fase de gerenciamento do sistema deveria ser parte integral de todo o projeto de construção do vocabulário controlado a fim de que o mesmo possa ser finalizado com êxito. Algumas características desejadas, segundo a norma, para os sistemas de gerenciamento de vocabulários controlados são listadas abaixo:

- Suporte mínimo aos relacionamentos básicos de Use/Usado para, Termo Genérico/Termo Específico e Termo Relacionado.
- Suporte ao uso de notas de escopo e notas históricas.
- Fornecer amplas visualizações hierárquica e alfabética.
- De preferência ser de domínio público (não-proprietário).

- Ser independente de *hardware* e operar em sistemas de padrão aberto.
- Requerer pouco ou nenhum treinamento.
- Fornecer detalhes na documentação.

4.8.2 Pré-desenvolvimento

Em relação a um estudo de viabilidade, a norma ANSI/NISO Z39.19-2005 recomenda averiguar a existência de vocabulários controlados que cobririam o mesmo domínio de conhecimento ou um domínio similar antes da decisão de se construir um novo vocabulário controlado. Para tal, indica consultar a Bibliografia de listas de vocabulários controlados disponível na própria norma, além de portais web destinados ao assunto.

4.8.3 Especificação de Requisitos

A fase de especificação de requisitos não foi considerada explicitamente nos materiais consultados, entretanto o manual da BITI faz menção a algumas atividades que podem ser consideradas como pertencentes à fase de especificação de requisitos. A saber: i) *delimitação da área do assunto* a ser coberto pelo vocabulário especializado, de forma que os conceitos envolvidos sejam bem sistematizados; e ii) *definição do público-alvo* no qual o instrumento se destina, implicando em definições sobre o grau de especificidade e características conceituais na estrutura do vocabulário.

As tarefas de especificação de requisitos devem ser realizadas junto a especialistas do domínio, usuários que irão manusear o instrumento e especialistas de projetos de vocabulários controlados.

4.8.4 Modelagem Conceitual

A organização de um domínio de conhecimento requer uma atividade de modelagem a fim de conceber um modelo ou uma representação normativa que prediga as ocorrências em um campo de assunto. Nesse sentido, para a construção de um vocabulário controlado de determinado assunto, um modelo conceitual é construído através dos seguintes passos[132]: i) seleção de termos representativos do assunto; ii) definição dos termos; iii) agrupamento dos termos em categorias; e

---

[132] Os passos foram determinados a partir da análise das recomendações feitas na norma e dos procedimentos encontrados no manual da BITI. Entretanto, a atividade de modelagem não é mencionada explicitamente em tais materiais. Desse modo, tornou-se possível inferir os passos a partir do conhecimento prévio adquirido na pesquisa.

iv) organização das relações. As seções a seguir apresentam tais passos de acordo com a metodologia proposta no manual da BITI e recomendações da norma ANSI/NISO Z39.19-2005.

4.8.4.1 Seleção de termos

A seleção de termos é um dos fatores mais importantes na criação de um vocabulário controlado, pois implica diretamente na aceitação do produto pelos usuários, além de afetar principalmente os indexadores e os pesquisadores na utilização dos termos.

O passo vislumbra identificar os termos representativos de conteúdo dos objetos pertencentes a um domínio de conhecimento, a fim de serem incluídos no vocabulário controlado. Segundo a metodologia disposta no manual da BITI, tais termos seriam direcionados a uma **planilha de registro dos dados** (ver detalhes no sítio inerente ao manual), contendo os seguintes campos: *assunto do vocabulário*, *categoria* (identificada pelo processo de categorização), *termo*, *definição*, *nota explicativa*, *termo(s) não-preferido(s)*, *termo(s) genérico(s)*, *termo(s) específico(s)*, *termo(s) partitivo(s)*, *termo(s) associado(s)* e *observações e indicação da(s) fonte(s)*. Ainda, segundo a metodologia, o critério de seleção dos termos tem por base a delimitação da área feita na fase de planejamento ou especificação de requisitos.

Segundo a norma ANSI/NISO Z39.19-2005, a atividade de identificação envolve **analisar** o domínio através de consultas a várias fontes de conhecimento (levantadas na fase de planejamento), bem como **seguir** critérios baseados na: a) linguagem natural usada para descrever conteúdo de objetos, denominado *garantia literária*; b) linguagem dos usuários, denominado *garantia de uso*; e c) necessidade e prioridade de organização, denominado *garantia estrutural*. No caso da *garantia literária*, é indicado revisar objetos primários e secundários que o vocabulário deverá tratar na indexação, bem como consultar fontes de referência como dicionários e vocabulários existentes, que representam o conteúdo do domínio. Para a determinação da *garantia estrutural* seria necessária a identificação das formas dos termos preferidos no vocabulário. E, finalmente, a *garantia de uso*, que reflete o uso dos termos na requisição da informação ou através da busca de termos pelos usuários em um sistema de recuperação da informação.

A ANSI/NISO Z39.19-2005 recomenda dois métodos para identificação dos termos do domínio: o *top-down* e o *bottom-up*. Na primeira abordagem, os termos mais gerais são identificados e posteriormente especializados. A norma ressalta que tal abordagem deveria ser usada quando um novo vocabulário estiver sendo criado. A segunda abordagem ocorre quando listas de termos são derivadas de um conjunto de objetos já existentes e posteriormente os termos seriam movidos para um escopo mais genérico. Já a metodologia propõe o método de *exercício prévio de classificação*, de forma a estabelecer os limites do tema a ser coberto. Utiliza-se para isso

a técnica de categorização para pensar o domínio de forma dedutiva, determinando as classes de maior abrangência (facetas) dentro da temática escolhida. A análise de domínio segundo a categorização serve para orientar os profissionais no levantamento e seleção dos termos que devem integrar o instrumento.

4.8.4.2 Definição dos termos

O termo é a designação do conceito. Para tal, o passo de definição dos termos implica em definir o significado dos termos. Da mesma forma, como foi procedida na seleção dos termos, a definição também deve ser incluída na **planilha de registro dos dados**, segundo o manual da BITI. Tal manual tem como fundamentação para a definição de um conceito o triângulo conceitual de Dahlberg, conforme foi elucidado na seção 2.2.3.2.

A ANSI/NISO Z39.19-2005 recomenda[133], dentre uma série de tratamentos: a) eliminar ambigüidades; b) controlar termos sinônimos e quase sinônimos[134]; c) distinguir homógrafos; d) usar notas de escopo e notas históricas. Tais recomendações são importantes para o projeto e desenvolvimento do vocabulário controlado, e seus princípios são elucidados a seguir:

- Ambigüidades: assegurar que cada termo tenha apenas um significado.
- Sinônimos: assegurar que cada conceito seja representado por um único termo de preferência; o vocabulário deveria listar os outros termos sinônimos como não-preferidos com referências USE para os termos preferidos.
- Homógrafos: clarear o significado do termo através de qualificadores, que são palavras envolvidas entre parênteses, usadas juntamente com o termo para torná-lo não-ambíguo.
- Notas de escopo: são usadas para restringir ou expandir a aplicação de um termo; distinguir entre termos que têm significados similares em linguagem natural ou fornecer outras informações sobre o termo usado.
- Notas históricas: são usadas para deixar registros acerca do desenvolvimento de um termo durante um período.

[133] Foram apresentadas algumas das recomendações, sendo que a norma recomenda uma série de tratamentos em relação ao termo, a saber: forma do termo, formas gramaticais dos termos, nomes usados como termos, seleção da forma preferida e termos compostos. Informações adicionais podem ser encontradas no capítulo 6 e 7 da norma.
[134] Diz-se que há quase-sinonímia quando dois conceitos têm praticamente a mesma intensão (CAMPOS, GOMES e MOTTA, 2004).

4.8.4.3 Agrupamento dos termos em categorias

Após, ou no momento da definição dos termos, o próximo passo seria sistematizar os conceitos do vocabulário controlado.

O método de categorização fornece os princípios para agrupar conceitos de mesma natureza em classes gerais ou facetas para construir cadeias e renques (CAMPOS, GOMES e MOTTA, 2004). O entendimento das classes que compõem um dado domínio é de suma importância para a elaboração de vocabulários controlados, pois permite uma maior compreensão do conceito e da organização das relações entre os conceitos.

Segundo o manual da BITI, o profissional envolvido na organização do domínio pode usar uma lista de categorias desenvolvida pelo CRG, a saber: i) *coisas, substâncias, entidades*; ii) *suas partes*; iii) *sistemas de coisas*; iv) *atributos de coisas*; v) *objeto da ação*; vi) *relações entre coisas*; vii) *operações sobre coisas*; viii) *propriedades de atributos*, *relações* e *operações*; ix) *lugar*, *condição*; x) *tempo*. Uma observação relevante sobre o processo de facetação é que nem todos os domínios possuem manifestações (facetas) para todas as categorias (CAMPOS, GOMES e MOTTA, 2004).

Posteriormente à definição das categorias, os termos que foram previamente selecionados são classificados nas respectivas facetas, as quais servirão de base para a formação das classes conceituais. A identificação das classes, no entanto, é feita a partir da literatura, isto é, deve ter garantia literária.

A ANSI/NISO Z39.19-2005 também indica o princípio da análise facetada como uma possibilidade de organizar conhecimento, e é particularmente usada para: a) novos e emergentes campos; b) áreas interdisciplinares; c) vocabulários em que múltiplas hierarquias são requeridas; e d) classificação eletrônica de documentos. Segundo a norma, a análise facetada é usada às vezes para indicar atributos de conteúdo de objetos, facilitando a localização de itens nas buscas. Os atributos que podem ser selecionados como facetas para conteúdo de objetos são:

- Tópico: o assunto do conteúdo do objeto.
- Formato: o formato do material (texto, imagem, som, etc.).
- Público-alvo: o leitor apropriado para o conteúdo (criança, adulto).
- Nível intelectual: o tipo de conteúdo (notícias, publicação acadêmica, etc.)

4.8.4.4 Organização das relações

O passo de organização das relações pode ocorrer simultaneamente com os passos de definição e agrupamento dos termos, isto é, com tarefas ligadas à organização dos termos e conceitos.

O manual da BITI indica conhecer primeiramente a natureza dos relacionamentos para posteriormente empregar os tipos de relações pertinentes. As relações expostas no manual seguem uma linha terminológica (ver seção 2.1.2), sendo classificadas em relações **lógica** e **ontológica**. A relação lógica vai produzir a relação hierárquica, que reúne conceitos que têm características comuns entre si, formando cadeias (subordinação ou espécie; superordenação ou gênero) e renques (coordenação). A relação ontológica vai incluir relações partitivas e associativas. No que diz respeito à partitiva, a relação entre o todo e as partes pode ocorrer nas categorias Coisa, Atividade e Lugar, segundo o manual da BITI. Alguns exemplos são elucidados: i) objeto integral/componente: os componentes possuem uma funcionalidade específica e podem ser separáveis, como roda sendo parte de um carro; ii) membro/coleção: os membros fazem parte do todo, e tais membros podem ser individualizados, como é o caso de uma árvore sendo parte de uma floresta; iii) objeto/matéria: a parte não pode ser separada do objeto constitutivo, pois não tem função sem tal objeto, como o aço (material) fazendo parte da bicicleta (objeto); iv) atividade/etapa: a etapa estabelece um papel funcional com a atividade, não podendo ser separada da mesma, como a catalogação, que pertence a atividade de tratamento de documentos; e v) área/lugar: relação espacial em uma região ocupada por diferentes objetos, como oásis fazendo parte de um deserto. No caso da associativa, a relação acontece por contigüidade. O manual exemplifica os tipos mais freqüentes: i) material/produto: o couro dando origem ao calçado; ii) processo/agente: processamento de dados/computador; iii) ação/resultado da ação: indexação/índice; iv) causa/efeito: chuva/acidente de carro; e v) processo/etapas de um processo: ciclo de vida/ infância, adolescência, juventude, maturidade e velhice.

A norma ANSI/NISO Z39.19-2005 classifica as relações em três tipos: relação de equivalência, hierárquica e associativa. A relação de equivalência é dividida em: relações de sinonímia, variações lexicais entre conceitos e quase-sinonímia. Tais relações representam os sinônimos ou quase sinônimos de um termo. Os sinônimos aparecem com a indicação de USE ou UP (use para), nos tesauros, mas não há distinção sintática para sinônimos e quase-sinônimos. A relação hierárquica engloba a relação do tipo genérica (gênero-espécie); instância (é um) e relação do tipo partitiva (todo-parte), ou seja, o todo com suas partes e do tipo enumerativa e as poli - hierárquicas, onde um termo pode depender de mais de um termo genérico. Finalmente, a relação associativa pode ocorrer como causa/efeito, processo/agente, processo/contra-agente, ação/produto,

ação/propriedade, ação/alvo, conceito ou objeto/propriedade, conceito ou objeto/origem, conceito ou objeto/ unidade ou mecanismo de medida, matéria bruta/produto e disciplina ou campo/objeto ou profissional. As relações associativas são estabelecidas entre termos que não são enquadrados como relação hierárquica ou equivalente.

O manual da BITI indica incluir as relações na planilha de registro dos dados à medida que a organização acontece. E, ainda, que o sistema de conceitos deve ser representado em gráficos até se chegar a uma solução definitiva.

4.8.5 Formalização

Após o modelo ter sido concebido, a próxima atividade seria direcionada à formalização ou à apresentação do vocabulário controlado. Tal atividade implica diretamente na satisfação dos usuários em relação ao manuseio do instrumento.

O manual da BITI indica uma série de representações gráficas[135] para as relações (expostas na seção 4.8.4.4), nas quais irão auxiliar na visualização do sistema de conceitos quando da evidência das relações presentes. Tais representações irão interferir no ato de modelar, pois os gráficos precisarão se adequar às relações concebidas na modelagem conceitual do vocabulário. A formalização recomendada pelo manual da BITI para representação gráfica de tesauro seria: i) para a relação hierárquica: lista estruturada, em árvore e combinação de gráficos; ii) para a relação partitiva: lista estruturada, em chave e combinação de gráficos; e iii) para a relação associativa: seta bidirecionada.

Já a ANSI/NISO Z39.19-2005 recomenda tratar os seguintes elementos na formalização:

- Apresentação: as decisões de apresentação[136] incluem como representar os relacionamentos dos termos.
- Tipos de exibição: podem ser desde listagens alfabéticas simples até visualizações gráficas complexas (também chamadas de apresentações sistemáticas).
- Formato: os vocabulários controlados podem ser distribuídos em formatos impressos ou eletrônicos (recursos de hipertexto).
- Documentação: todo vocabulário controlado deveria exibir uma documentação de como usar o instrumento.

[135] Exemplos ilustrativos de tais gráficos podem ser encontrados no referido manual.
[136] Informações adicionais podem ser encontradas na seção 9.2 da norma.

- Categorias de usuários: verificar as classes de usuários que irão manusear o instrumento a fim de fazer um estudo de usabilidade para cada classe, como indexadores, usuários finais e pessoal de apoio (manutenção do instrumento).

4.8.6 Manutenção

Essa fase é destinada a procedimentos relacionados à adição, à alteração e à exclusão de termos no vocabulário controlado. Para tal, a norma ANSI/NISO Z39.19-2005 recomenda fazer o registro do termo que sofreu a manutenção, incluindo a data do evento e o responsável pela manutenção. As seguintes recomendações são direcionadas a cada tipo de manutenção:

- Adição: verificar a necessidade de incluir notas de escopo para o novo termo; verificar qual é a forma correta do termo; verificar como o novo termo será relacionado com os termos existentes na hierarquia.
- Alteração: registrar a data da alteração dos termos na nota histórica. Isso procede também quando os relacionamentos forem alterados.
- Exclusão: quando excluído, o termo pode ser mantido no vocabulário apenas como propósito histórico e deve ser marcado como "apenas como propósito histórico", além da data ser registrada na nota histórica do termo.

O manual da BITI ressalta que um grupo deve ser criado para a atividade de manutenção do tesauro, e que as manutenções, no caso específico de uma inclusão de novos termos, devem ser registradas em uma planilha específica contendo os seguintes campos: termo, categoria, definição, termo equivalente, termo genérico, termos específicos e termos associados. Ressalta-se que tais relações entre termos devem ser mencionadas quando da pertinência ao contexto do domínio.

4.8.7 Avaliação

Segundo a norma ANSI/NISO Z39.19-2005 , a fase de avaliação ou de testes é relevante para determinar se o vocabulário controlado está fornecendo resultados de busca satisfatórios, os quais implicam em uma boa relação entre precisão e revocação.

Uma das tarefas da avaliação, de acordo com a norma, seria validar os termos, considerando o seu escopo (significados) e a sua forma (relacionamentos entre cada termo), antes de serem direcionados ao vocabulário controlado. Vários critérios de avaliação de vocabulários controlados em formato de questões são propostos na norma ANSI/NISO Z39.19-2005 e podem ser consultados na seção 11.2.2 da mesma. A norma ainda recomenda ainda executar testes de usabilidade sobre o

vocabulário controlado. Recomendações sobre testes de usabilidade podem ser encontradas na literatura sobre interface homem-máquina e arquitetura de informação, segundo a norma.

O manual da BITI recomenda alguns critérios para fundamentar decisões quanto à adoção ou não de um tesauro existente em um ambiente de serviço de informação. São eles:

1. Domínio de conhecimento coberto pelo tesauro.
2. Apresenta introdução? Qual o conteúdo? Está redigida de forma clara a possibilitar o uso do instrumento?
3. Forma de apresentação
   a) Apresenta parte alfabética
   b) Apresenta parte sistemática
4. Idioma: monolíngües ou multilíngüe?
5. Unidade lingüística utilizada:
   a) Conceito
   b) Palavra
   c) Assunto
6. Quais os tipos de relação encontrados?
7. Aspectos ligados à consistência:
   a) Consistência das relações entre os termos?
   b) Consistência no uso do plural e do singular?
   c) Consistência no nível de especificidade?
8. Nota de aplicação/Escopo: apresenta a definição do termo e/ou a política de indexação?

4.8.8 Documentação

A norma ANSI/NISO Z39.19-2005 recomenda que quando o vocabulário encontra-se no formato impresso, a documentação deveria aparecer na introdução. Entretanto, se o formato estiver disponível *on-line*, a documentação também deveria estar disponível *on-line* (através de arquivo PDF ou no contexto de ajuda *on-line*). Algumas recomendações da norma, direcionadas ao conteúdo da documentação, são listadas abaixo:

- O propósito do vocabulário controlado.
- O escopo, ou seja, o campo coberto.
- O significado de todas as abreviaturas e convenções.
- As regras adotadas na seleção das formas preferidas dos termos e o estabelecimento de seus relacionamentos.
- Se o vocabulário compreende padrão nacional ou internacional.
- Regras empregadas através de guias ou padrões recomendados.
- O número total de termos com totais separados para termos e entrada de termos.
- A data da última atualização do vocabulário controlado.

- A declaração da política de atualização.
- A informação de contato do responsável pelo vocabulário, de forma que comentários e sugestões possam ser enviados.
- Opções de navegação especial *on-line* quando disponível.

Já o manual da BITI fornece detalhes de como documentar a construção de um tesauro em uma planilha de registro dos dados (elucidada na seção 4.8.4.1). A manutenção de um tesauro também seria documentada em uma planilha específica, porém o manual não fornece detalhes referentes aos campos a serem documentados. Tais planilhas irão possibilitar o registro do conhecimento sobre a concepção e a evolução do instrumento.

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Deveria ser **parte integral de todo o projeto** de construção do vocabulário controlado. Recomenda características necessárias aos sistemas de gerenciamento de vocabulários controlados.	Recomenda um **estudo de viabilidade** para a construção.	Ausente	Recomenda **tratamentos em relação ao termo**; organização das **relações** em **hierárquica**, **associativa** e de **equivalência**.	Recomenda tratar a **representação dos relacionamentos dos termos**.	Ausente	Recomenda procedimentos para a **adição**, a **alteração** e a **exclusão** de termos.	Ausente	Determina se o vocabulário controlado está fornecendo **resultados de busca satisfatórios**, que implica em uma boa relação entre precisão e revocação. Uma das tarefas da avaliação seria **validar os termos** antes de serem direcionados ao vocabulário controlado.	Recomendações quanto ao **conteúdo** da **documentação**.

Tabela 21 – Tabela sinóptica da norma ANSI/NISO Z39.19-2005

Gerenciamento do projeto	Processos orientados						Processos integrais		
	Pré-desenvolvimento	Desenvolvimento				Pós-desenvolvimento			
		Especificação de requisitos	Modelagem conceitual	Formalização	Implementação	Manutenção	Integração	Avaliação	Documentação
Atividades de planejamento do projeto.	Ausente	Planejamento prévio à construção: i) **delimitação** da **área**; e ii) definição do **público-alvo.**	i) **seleção** de termos representativos do assunto; ii) **definição** dos termos; iii) **agrupamento** dos termos em **categorias**; e iv) **organização** das **relações,** que são classificadas em **relação lógica e ontológica**.	**Tipos de exibição** do tesauro, incluindo **listas alfabéticas** e **visualizações gráficas**, formatos impressos ou eletrônicos.	Ausente	A **inclusão de novos termos** deve ser **proposta** numa **planilha específica**.	Ausente	**Critérios** adotados para a **adoção** ou **não** de um **tesauro** existente.	Registro dos dados em **planilhas**, que irão possibilitar o registro do conhecimen-to sobre a **concepção** e a **evolução** do instrumento.

Tabela 22 – Tabela sinóptica da metodologia proposta no manual da BITI

5 - Apresentação da Análise Comparativa

O presente capítulo apresenta a análise comparativa das metodologias e dos métodos para construção de ontologias, e da metodologia e da norma para construção de vocabulários controlados.

Para a execução da análise comparativa, foi necessário um entendimento dos procedimentos metodológicos dos objetos investigados. Tal entendimento tornou-se possível a partir da análise de conteúdo feita nos materiais empíricos, conforme apresentada no capítulo 4 desta pesquisa. Além disso, para que as discussões pudessem ser feitas nesse capítulo, foi necessária uma compreensão ampla sobre diversos conceitos que circundam a temática principal deste estudo, as *ontologias*, bem como das teorias advindas dos campos das Ciências da Informação e da Computação.

A análise comparativa foi realizada a partir de cada categoria denominada, a partir desse ponto, *fase do ciclo de vida* envolvida nos objetos investigados. Posto isto, o capítulo encontra-se organizado em seções que estabelecem discussões comparativas acerca de cada fase, a saber: *gerenciamento do projeto*; *pré-desenvolvimento*; *especificação de requisitos*; *modelagem conceitual*; *formalização*; *implementação*; *manutenção*; *integração*; *avaliação*; e *documentação*. A última seção finaliza o capítulo estabelecendo considerações gerais sobre a análise comparativa.

Na seqüência, uma tabela sinóptica representando a análise comparativa é apresentada, de forma a proporcionar uma visão abrangente do conteúdo de cada objeto investigado. Tal tabela foi baseada no trabalho de Fernandez, Gomez e Corcho (2004), no qual executam análises e comparações acerca de metodologias e métodos para construção de ontologias. As fases que **não foram propostas** nos objetos investigados tiveram suas células preenchidas com “Ausente”; quando contempladas **em detalhe**, a célula foi preenchida com “Descrito em detalhe”; e quando propostas, mas **sem muito detalhe**, a célula foi preenchida com “Proposto”.

Finalmente, uma segunda tabela é apresentada mostrando o mapeamento feito, a partir da análise comparativa, em métodos, técnicas e princípios que poderiam vir a ser úteis ao objetivo geral desta pesquisa. A célula preenchida com “Ausente” indicou a ausência de características que poderiam interessar à etapa final da pesquisa.

5.1 Gerenciamento do projeto

Na fase de *gerenciamento do projeto*, atividades relacionadas ao início, ao planejamento e ao monitoramento e controle do projeto são conduzidas em todo o ciclo de vida da ontologia.

Dos objetos investigados, apenas a **Methontology**, a **metodologia** para construção de **tesauros** e a **norma** para construção de **vocabulários controlados** contemplaram a fase de gerenciamento do projeto. Um ponto em comum entre as três abordagens está na consideração do gerenciamento durante todo o ciclo de atividades envolvido na construção do instrumento. A metodologia para construção de **tesauros** destaca atividades de planejamento do projeto como levantamento das fontes de pesquisa, decisão da forma de apresentação, período de atualização, seleção do *software* para construção do instrumento e definição de grupos para atuar em possíveis manutenções. Um ponto em destaque na **Methontology** é o planejamento de escalonamento, que propõe iniciar cada protótipo (ciclo de vida da ontologia baseado na evolução de protótipos) com uma escala de atividades, incluindo suas respectivas tarefas, arranjos, pessoas e recursos. Cada escala seria gerenciada por um controle e garantia de qualidade de forma a assegurar que a ontologia, a aplicação e a documentação sejam construídas com êxito. Além disso, a **Methontology** e a **ANSI/NISO Z39.19-2005** consideram a fase de gerenciamento do projeto como sendo estratégica para a garantia da qualidade do produto final, sendo, portanto, um ponto positivo em ambas as abordagens. Um ponto negativo é que a norma para construção de vocabulários controlados e a metodologia para construção de ontologias não fornecem detalhes de como proceder nas atividades envolvidas.

5.2 Pré-desenvolvimento

A fase *pré-desenvolvimento* consiste em estudar o ambiente atual no intuito de identificar as plataformas (*software*, *hardware*) e as aplicações em que a ontologia será usada e integrada. A fase também inclui atividades voltadas a estudo de viabilidade, no sentido de verificar questões a respeito da possibilidade e adequação de se construir ontologias em determinado ambiente.

Do universo de nove objetos investigados, apenas dois apresentaram uma proposta na fase pré-desenvolvimento: a metodologia de **Gruninger e Fox** e a norma **ANSI/NISO Z39.19-2005** para construção de vocabulários controlados. A metodologia proposta no manual da **BITI** não menciona atividades relacionadas à fase pré-desenvolvimento. As demais metodologias e métodos para construção de ontologias mostram não se preocupar com tal fase, partindo, na maioria dos casos, diretamente para a fase de especificação de requisitos da ontologia.

É válido destacar que no processo de análise da metodologia de **Gruninger e Fox** surgiu uma dúvida em relação à devida classificação da atividade *cenários de motivação*, visto que tal atividade, além de servir para descrever os requisitos da ontologia, também tem o propósito de identificar problemas que podem refletir na decisão de construção de novas ontologias ou extensão de ontologias já existentes, conforme foi descrito na seção sobre tal metodologia. O propósito de

identificar problemas em um ambiente se enquadra em atividades pertencentes à fase pré-desenvolvimento, conforme foi visto na fundamentação teórica sobre os grupos de atividades da norma IEEE-1074-1997. Portanto, achou-se pertinente classificar os cenários de motivação na categoria pré-desenvolvimento.

No caso da **ANSI/NISO Z39.19-2005**, a norma recomenda um estudo de viabilidade antes da decisão de se construir um vocabulário controlado, porém, não entra em detalhes sobre a forma de condução de tal estudo.

A fase pré-desenvolvimento mostra-se relevante no sentido de averiguar a razão de se construir ontologias em determinado ambiente e pode implicar, muitas vezes, em propostas de desenvolvimento, as quais necessitam ser bem definidas e detalhadas a fim de que o projeto de ontologias possa ser conduzido de forma coerente junto a tais propostas.

5.3 Especificação de Requisitos

A fase de *especificação de requisitos*, pertencente ao processo de desenvolvimento da ontologia, permite identificar o propósito de construção e os usuários finais da ontologia. Para tal, faz-se necessário adquirir conhecimento sobre os problemas que circundam o ambiente, de forma a especificá-los através de questões que a ontologia deveria ser capaz de solucionar. A atividade de aquisição de conhecimento é executada junto a um especialista do domínio a fim de que os requisitos especificados pelo *ontologista* sejam validados pela pessoa entendedora do domínio.

Na análise executada sobre as metodologias e os métodos para construção de ontologias, apenas o método **Sensus** não apresentou uma proposta de especificação de requisitos, dando indícios de que o início do processo de construção de ontologias concentra-se na fase de modelagem conceitual. Tal procedimento não se mostra muito correto, visto que a atividade de especificação de requisitos permite ao *ontologista* entender a *necessidade* de se construir ontologias. Para as demais metodologias e métodos houve a preocupação em determinar o assunto ou o domínio a ser coberto, os tipos de usuários e a intenção do uso da ontologia. Acredita-se que somente com essas informações disponíveis um modelo conceitual poderá ser bem concebido, pois a validação do mesmo poderá ser feita a partir do seu confrontamento com as necessidades elucidadas na especificação.

No contexto dos vocabulários controlados, a norma **ANSI/NISO Z39.19-2005** não faz menção à fase de especificação de requisitos. Já o manual da **BITI** considera atividades que contemplam tal fase: delimitação da área e definição do público que irá manusear o tesauro. Tais atividades irão implicar nos requisitos necessários à construção do tesauro.

No contexto das ontologias, destaca-se a especificação de requisitos elaborada através das *questões de competência informal*. Tais questões permitem ao *ontologista* construir modelos mais próximos dos problemas elucidados a partir dos cenários de motivação e avaliar a expressividade da ontologia a partir da sua competência em responder às questões. Dentre os objetos investigados, a metodologia de **Gruninger e Fox** foi a que mais se destacou quanto ao método de especificação através de questões de competência. As razões seriam: i) originalidade; ii) descrição de cenários de motivação que implicam nos requisitos da ontologia; iii) preocupação em explicar a importância do método; e iv) apresentação de exemplos ilustrativos que facilitam o entendimento do método. O método de **Uschold e King** e a metodologia **Methontology** apenas mencionam a possibilidade de utilizar tal abordagem. Já o **método 101** apresenta alguns exemplos de questões de competência e algumas orientações para a elaboração das mesmas.

Dentre as metodologias e métodos investigados, apenas a **Methontology** considerou a atividade de identificar os termos relevantes (chamado de escopo) para as ontologias na fase de especificação de requisitos. Entretanto, tais termos seriam posteriormente refinados na fase de modelagem conceitual. A atividade de *aquisição de conhecimento* é bem destacada na fase de especificação da metodologia, indicando possíveis fontes de conhecimento a consultar e técnicas para se obter conhecimento, destacando *brainstorming*, entrevistas estruturadas, análise informal de documentos e métodos de extração de conhecimento com o auxílio de ferramentas computacionais. O método proposto na metodologia para a identificação de termos é o *middle-out* que vai de encontro à visão de **Uschold e King** de que o método proporciona menos esforço e re-trabalho, além de tornar os conceitos mais estáveis.

5.4 Modelagem Conceitual

Nessa pesquisa, considera-se a fase de *modelagem conceitual* como uma das mais importantes no processo de desenvolvimento de ontologias, pois possibilita organizar e representar de forma coerente os elementos de um determinado domínio, incluindo conceitos, atributos, relações, instâncias, constantes e termos. É na fase de modelagem que se torna possível determinar um conjunto de compromissos ontológicos fundamentais para a construção da conceitualização de senso comum, que poderá ser compartilhada em um domínio de interesse.

Na concepção da conceitualização, isto é, do conjunto de modelos pretendidos (GUARINO, 1998), algumas metodologias se mostraram mais preocupadas em detalhar o processo do que outras. Ocorreu ainda de a conceitualização ser desconsiderada ou pouco considerada dentro do processo de construção da ontologia. Apenas o método **Cyc** não considerou a fase de modelagem conceitual, partindo diretamente para a codificação do conhecimento extraído manualmente e automaticamente

(com apoio de ferramentas computacionais) das fontes. O método **Cyc** parece não considerar muito relevante o trabalho de ordem intelectual, isto é, o mecanismo de abstração semântica por parte da pessoa que modela domínios de conhecimento. Nos modelos conceituais, conforme assinala Sayão (2001), as abstrações semânticas tornam-se relevantes por considerar relações entre conceitos lingüísticos que refletem diferenças de significados entre termos. Tais abstrações apenas conseguem atingir um grau aceitável de aproximação com o objeto referente a partir de um trabalho intelectual por parte do profissional que modela o sistema. O método **Cyc** associa completamente a atividade de organização do conhecimento a técnicas computacionais, mostrando-se vulnerável em relação à conceitualização da ontologia. A **Methontology** apresenta detalhes do processo de construção da conceitualização através das chamadas *representações intermediárias*, destacando as *árvores de classificação de conceitos, de atributos* e *de verbos*. Tais representações mostram-se úteis no sentido de possibilitar uma construção metódica do conhecimento a partir de um consenso. Nessa pesquisa, a **Methontology** é considerada uma referência em termos de procedimentos de modelagem que buscam organizar um domínio a partir do nível de conhecimento. A metodologia de **Gruninger e Fox** também se ocupou em deixar clara a representação da terminologia ou taxonomia através da lógica de primeira ordem, mostrando-se, assim como a **Methontology**, eficiente quanto aos métodos elucidados na metodologia. O método de **Uschold e King**, apesar de indicar métodos e técnicas para identificação de conceitos (sessão de *brainstorming*, abordagens *bottom-up*, *top-down* e *middle-out*) e algumas orientações quanto ao tratamento de termos sinônimos e ambíguos, não expõe os procedimentos necessários à concepção de uma conceitualização. Tal constatação vai de encontro à visão de Gómez-Perez, Fernandez e Corcho (2004), de que a principal deficiência do método de **Uschold e King** é a falta do processo de conceitualização antes da implementação da ontologia. A concepção da conceitualização através do **método 101** mostrou ser um processo altamente dependente de tecnologia, isto é, as atividades de identificação de termos e suas propriedades, a definição de classes, a definição de propriedades de classes, a definição de restrições de classes e o tratamento de instâncias são conduzidas através de uma ferramenta de edição de ontologias. Sendo assim, o método não se mostrou preocupado com questões de projeto, partindo direto para a fase de formalização e implementação da ontologia. Por outro lado, o **método 101** contribui no aspecto tecnológico, visto que apresenta em detalhe e com clareza os passos para se construir uma ontologia em uma das *ferramentas mais utilizadas* (CARDOSO, 2007, p.4) por desenvolvedores de ontologias, o *Protégé*. Além disso, o método também elucida um conjunto de princípios que parece ser interessante no momento de se definir as classes e estruturá-las em uma taxonomia. Já os métodos **Kactus** e **Sensus** se mostraram superficiais em relação à fase de modelagem, não abordando com clareza os procedimentos necessários à concepção da conceitualização. Apenas ressaltam a importância do reuso de conhecimento a partir de ontologias

de alto nível para a construção de ontologias de *aplicação* e de *domínio*, respectivamente no caso do método Kactus e do método Sensus.

Finalmente, a metodologia proposta no manual da **BITI** e a norma para construção de **vocabulários controlados** investigadas, apesar de não nomearem a fase em questão de modelagem conceitual (geralmente chamada de estruturação do domínio) em seus procedimentos e recomendações para construção, buscam organizar os elementos de um domínio através de decisões intelectuais[137] associadas à seleção de termos representativos, à definição precisa de termos, ao agrupamento de termos em categorias e à organização de relações. Um recurso tecnológico poderia ser utilizado posteriormente para formalizar ou apresentar o vocabulário. Além disso, o manual da **BITI** apresenta teorias que fundamentam seus princípios metodológicos, como as teorias da classificação facetada e do conceito. Desse modo, acredita-se que alguns métodos, técnicas e recomendações advindas da metodologia e da norma possam vir a ser úteis no processo de construção de ontologias, conforme se pretende elucidar no capítulo 6 deste estudo.

5.5 Formalização

Na fase de *formalização*, também pertencente ao processo de desenvolvimento da ontologia, o modelo conceitual é convertido em um modelo formal usando uma linguagem de representação de conhecimento como *frames*, redes semânticas ou declarações lógicas, conforme apresentados na seção 2.3.1 dessa pesquisa.

A metodologia de **Gruninger e Fox** se destaca por apresentar detalhes sobre os procedimentos de representação formal dos elementos constitutivos das ontologias envolvidas. A metodologia contempla a representação formal das questões de competência informal em coerência à taxonomia concebida na fase de modelagem, e a especificação dos axiomas formais para expressar as questões de competência formal e caracterizar as suas soluções. Os métodos **Cyc**, de **Uschold e King** e **Kactus** não apresentaram propostas de formalização. A **Methontology** não considera obrigatória a fase de formalização, pois alega que as próprias ferramentas utilizadas pela metodologia já tratam das questões de representação formal. O método **Sensus** apenas menciona o formalismo utilizado em seus construtos (rede semântica) sem entrar em detalhe sobre o processo de formalização da rede de conceitos através da ferramenta OntoSaurus. E, finalmente, o **método 101** realiza a formalização concomitante à fase de modelagem, conforme discutido na seção anterior. Para tal, utiliza uma linguagem de representação baseada em *frames*.

[137] Assistência por máquina pode ser empregada na identificação de termos candidatos e no registro de freqüência de termos selecionados (ANSI, 2005, p.92).

A especificação de ontologias através de um formalismo lógico torna-se vantajosa por possibilitar validações precisas em sua estrutura, verificando, assim, se tal estrutura corresponde ao conhecimento adquirido sobre o domínio. Uma outra vantagem seria na possibilidade de averiguar se as primitivas utilizadas em um determinado formalismo lógico são compatíveis com a linguagem que será usada na implementação de ontologias, pois alguns problemas (representações) resolvidos através de uma linguagem formal são computacionalmente intratáveis. Conforme assinalado em Rezende (2003), a complexidade de processamento pela máquina fica explícita na formalização dos conceitos. Dessa forma, o *ontologista* precisa saber escolher a linguagem de representação que melhor se enquadre nas necessidades da ontologia.

No contexto dos **vocabulários controlados**, a formalização não tem o mesmo propósito do que nas ontologias, ou seja, não contempla nenhum formalismo lógico na representação do vocabulário. A formalização abrange os tipos possíveis de exibição (lista alfabética, gráficos) e formato (impressos, eletrônicos) dos vocabulários controlados destinados a seus usuários.

5.6 Implementação

A *implementação*, última fase do processo de desenvolvimento, é contemplada em praticamente todas as metodologias e métodos para construção de ontologias investigados nessa pesquisa. Tal constatação se deve ao fato de que ontologias são especificadas para servirem de modelos computáveis aos sistemas de informações (GUARINO, 1998; SMITH, 2004). Desse modo, para que os modelos sejam especificados através de ontologias, faz-se necessária uma linguagem legível para a máquina que seja capaz de atender aos compromissos ontológicos estabelecidos com a conceitualização da ontologia, conforme foi elucidado na seção 2.3.3.1 da fundamentação teórica desta pesquisa. Para tal, a linguagem deve considerar axiomas formais em suas primitivas, de forma a fornecer a definição semântica do domínio considerado, restringindo, assim, o vocabulário utilizado.

Apenas o método **Kactus** não menciona procedimentos de implementação, dando ênfase à construção de modelos preliminares de aplicações específicas, os quais seriam mais tarde integrados a outras ontologias de aplicação. Entretanto, não deixa claro se tal integração é feita através de alguma ferramenta ou linguagem de implementação de ontologias.

O método de **Uschold e King** se destaca na fase de implementação por fazer menção a linguagens geralmente usadas na implementação de ontologias e considerar que a experiência no uso pode servir de ponto de partida para o desenvolvimento de um guia focado em determinar linguagens apropriadas para projetos específicos. Além disso, o método se mostra coerente com a fase de implementação quando estabelece critérios para escolher a linguagem de implementação de

ontologias. O mesmo acontece com a **Methontology** quando propõe critérios para a escolha de ambientes para desenvolvimento de ontologias. O **método 101** também se mostra relevante no sentido de permitir o manuseio interativo[138] com o código da linguagem escolhida para implementação através da ferramenta *Protégé*.

A metodologia de **Gruninger e Fox** e os métodos **Cyc** e **Sensus** mencionam as linguagens utilizadas na implementação de suas ontologias, porém não fornecem detalhes sobre o processo de implementação.

No caso dos **vocabulários controlados**, a metodologia e a norma investigadas não fazem menção a parte de implementação ou geração de código, mesmo a norma considerando formatos de exibição emergentes, como o hipertexto. Em muitos casos, utilizam-se ferramentas ou *softwares* para a construção de vocabulários controlados, mas apenas como apoio ao processo de construção e disponibilização do instrumento ao usuário.

5.7 Manutenção

A fase de *manutenção* pertence ao processo pós-desenvolvimento da ontologia. A sua importância se dá no sentido de considerar possíveis alterações em função de eventuais problemas mapeados na fase de avaliação da ontologia, bem como em necessidades de inclusão, alteração e exclusão de conceitos, termos, enfim, elementos importantes da ontologia. Nesse contexto, somente a **Methontology** considera tal fase como sendo importante no processo de construção. A ausência dessa fase coloca em evidência a vulnerabilidade das outras metodologias e métodos investigados frente a tratamentos necessários e corretos nos elementos que sofrem impacto diante das manutenções. A **Methontology** recomenda verificações e ajustes nas representações intermediárias (ver tabela 16), caso ocorram modificações em elementos constitutivos das mesmas em função de alterações nos protótipos envolvidos no processo de construção de ontologias.

Já no contexto dos vocabulários controlados, a norma **ANSI/NISO Z39.19-2005** recomenda tratamentos na estrutura do vocabulário controlado quando da adição, alteração e exclusão de termos, evitando possíveis inconsistências em função de eventuais modificações. E o manual da **BITI** considera o uso de uma planilha específica a fim de registrar as manutenções efetuadas, entretanto não apresenta detalhes de como proceder na elaboração de tal planilha.

[138] O *Protégé* permite a especificação de ontologias através de um *framework*, o qual facilita a atividade de desenvolvimento da ontologia por não exigir muito trabalho em termos de codificação. Por possibilitar a geração automática de código, o desenvolvedor consegue se interagir mais facilmente com o código através de recursos disponíveis na ferramenta.

5.8 Integração

A fase de *integração* faz parte dos processos integrais, ou seja, processos executados ao mesmo tempo com os processos orientados a desenvolvimento, e sem os quais a ontologia não poderia ser construída. No caso específico da integração, suas atividades podem ser conduzidas nas fases de modelagem conceitual, implementação e manutenção.

Ontologias geralmente são construídas reusando outras ontologias já disponíveis, isto é, aproveitando conceitos de conceitualizações que conseguem se adequar à realidade da conceitualização que está sendo construída. Todas as metodologias e métodos para construção de ontologias analisados consideram a fase de integração, confirmando que a integração é parte importante do processo de construção de ontologias, dado que tal processo é considerado dispendioso na medida em que estruturas intensionais e extensionais precisam ser construídas. Desse modo, a integração fica como uma alternativa para facilitar a construção de ontologias.

O método **Cyc** considera a integração das diferentes micro-teorias existentes na *base de conhecimento Cyc* com ontologias de variadas complexidades como a SENSUS e a WordNet, conforme apresentado na seção 4.1.2. Para tal, utiliza padrões tecnológicos de indústria como DAML, XML *Schema* e UML para integrar sistemas computacionais à base Cyc. Apesar de o método recomendar e explicar tecnicamente tais padrões para o processo de integração, não explicita os procedimentos metodológicos para estabelecê-la.

A metodologia de **Gruninger e Fox** apenas recomenda formalizar as perspectivas particulares de cada empresa para então integrá-las a ontologias de núcleo comum (representam um conhecimento comum de uma organização), e adverte quanto aos axiomas especificados nas ontologias específicas, os quais devem ser suportados pelas ontologias de núcleo comum. Tal recomendação não se mostra suficiente em termos metodológicos para facilitar a construção de ontologias usando-se como alternativa a integração.

O método de **Uschold e King** também não apresenta os procedimentos necessários à integração com ontologias já existentes, apenas recomenda verificar tal possibilidade e, quando da decisão de integrar, explicitar todas as pretensões subjacentes à ontologia de forma a alcançar uma correspondência semântica com a ontologia usada na integração.

Apesar do processo de reutilização de ontologias desenvolvidas por outras aplicações ser destacado no método **Kactus**, o mesmo não apresenta os procedimentos metodológicos necessários ao refinamento e extensão dos conceitos reutilizados. O método, mais uma vez, se mostra superficial na elucidação de seus procedimentos. O mesmo cenário acontece com o método **Sensus**, cuja base é o reaproveitamento de conceitos de ontologias de alto nível. Porém, não demonstra clareza na explicação de seus passos metodológicos no que diz respeito à integração com tais

ontologias. O **método 101** considera a reutilização de ontologias, mas também não apresenta procedimentos para estabelecer a integração.

A única que apresenta e explica com clareza um método para estabelecer integração com meta-ontologias é a **Methontology**. Apesar de não detalhar tal método, a metodologia se mostra objetiva ao elucidar o procedimento que, no final, resulta em um documento de integração, que pode auxiliar o desenvolvedor no processo de integração, conforme apresentado na seção 4.5.6.

A **metodologia** para construção de **tesauros** e a **norma** para construção de **vocabulários controlados** não foram classificadas na fase de integração pelo fato da mesma ser uma característica específica do processo de construção de ontologias.

5.9 Avaliação

A fase de *avaliação* pertence aos processos integrais pelo fato de suas atividades serem conduzidas ao término de cada fase do processo de desenvolvimento da ontologia. A avaliação tem como propósito julgar tecnicamente as ontologias, suas aplicações e ambientes associados, bem como a documentação desenvolvida durante o processo de construção.

Dentre as metodologias e métodos para construção de ontologias analisados, os métodos **Cyc**, **Kactus**, **Sensus** e **101** não contemplam a fase de avaliação, desconsiderando a importância da verificação sobre os resultados de cada fase do processo de construção, de forma que, no final, a ontologia possa ser concluída com a qualidade devida. E ainda desconsideram a relevância do procedimento de validação, a fim de averiguar se a ontologia está cumprindo o propósito especificado na fase de requisitos. Tais desconsiderações podem levar ao fracasso do projeto, visto que o sucesso do mesmo depende de homologações determinadas por verificações e validações em todo o ciclo de vida da ontologia.

A metodologia de **Gruninger e Fox** utiliza os teoremas completos para verificar se as soluções para as questões de competência são completas. Tal metodologia mostra ser madura no quesito avaliação, pois consegue validar com maior precisão ontologias a partir de seus axiomas formais. E ainda detalha os procedimentos necessários à especificação dos axiomas e aos testes de provação junto a teoremas. **Uschold e King** apenas recomendam utilizar questões de competência para avaliar ontologias, entretanto não apresentam procedimentos metodológicos necessários à avaliação. A **Methontology** ressalta a importância da fase de avaliação para o processo de construção, existindo propostas isoladas (GÓMEZ-PEREZ, 1994; GÓMEZ-PEREZ, 1999) para avaliação de ontologias, incluindo critérios de avaliação da taxonomia, dos conceitos e das definições que a compõe.

No contexto dos vocabulários controlados, a norma **ANSI/NISO Z39.19-2005** e a metodologia da **BITI** mostram-se eficazes no aspecto avaliação quando apresentam critérios de avaliação em formato de questões. Entretanto, as avaliações, em ambas as propostas, são direcionadas a processos diferentes. No contexto da **ANSI/NISO Z39.19-2005**, a avaliação é direcionada ao processo de recuperação de informação, que implica em uma boa relação de precisão e revocação; já o manual da **BITI** direciona a avaliação ao processo de adoção ou não de um tesauro existente.

5.10 Documentação

A fase de *documentação* também pertence aos processos integrais pelo fato de sua elaboração ocorrer ao longo de todo o ciclo de vida da ontologia. A documentação é extremamente relevante ao processo de construção da ontologia, pois detalha, clara e exaustivamente, toda a memória do projeto, incluindo informações acerca da metodologia adotada, dos requisitos levantados, do escopo, dos modelos, do código fonte, das manutenções ocorridas, das avaliações efetuadas, das integrações estabelecidas, enfim, tudo que se julga relevante em termos de gestão do conhecimento do projeto. Tal conhecimento pode ser utilizado para futuros ajustes que se fizerem necessários nas ontologias, bem como possíveis integrações, buscando manter, dessa forma, a consistência dos elementos pertencentes às ontologias envolvidas.

Dentre os objetos investigados para construção de ontologias, a metodologia de **Gruninger e Fox** não propõe documentações ao longo do ciclo de vida da ontologia, mostrando-se falha nessa parte. Os métodos **Cyc**, **Kactus** e **Sensus** também não contemplam a fase de documentação, desconsiderando a importância de se documentar os procedimentos de integração com ontologias de alto nível e aplicações diversas, visto serem métodos focados em reutilização de conhecimento. O método de **Uschold e King** considera a documentação somente na fase de especificação de requisitos, desconsiderando a sua relevância em outras fases do ciclo de vida. O **método 101** propõe documentações a partir da ferramenta *Protégé*, na qual permite descrever os objetos envolvidos na construção (classes, *slots*, facetas, instâncias, etc) em linguagem natural. O método também recomenda documentar premissas relevantes sobre o domínio como lista de sinônimos e limitações de escopo. A **Methontology** pode ser considerada a mais madura no aspecto documentação, principalmente nos procedimentos de elaboração das representações intermediárias, que servem de documentação para a fase de modelagem conceitual, bem como para a conceitualização. Tais documentos são descritos com clareza e exaustividade, conforme podem ser conferidos em Gómez-Perez, Fernandez e Vicente (1996). Além da fase de modelagem, a **Methontology** recomenda a

documentação como uma atividade a ser realizada durante todo o processo de desenvolvimento da ontologia, incluindo cada fase.

No caso dos vocabulários controlados, a **ANSI/NISO Z39.19-2005** recomenda uma documentação sobre a utilização do vocabulário e sobre a sua organização. Ainda apresenta um conjunto de recomendações sobre o conteúdo da documentação. O manual da **BITI** recomenda fazer o registro dos dados referentes à construção e à evolução de tesauros em planilhas específicas.

5.11 Considerações gerais sobre a análise comparativa

Diante das explanações realizadas nas seções do presente capítulo, chegou-se a algumas considerações sobre as metodologias e os métodos para construção de ontologias analisados e sobre a metodologia para construção de tesauros e a norma para construção de vocabulários controlados analisadas. Tais considerações são enumeradas a seguir:

- Existe uma variedade de estratégias para desenvolvimento de ontologias, comprovando a hipótese de que grupos diferentes apresentam abordagens e características diversas, sendo direcionadas a diferentes propósitos e aplicações (FERNÁNDEZ et al., 1999).
- No contexto das ontologias, algumas abordagens seguem um modelo de ciclo de vida, outras não. Nesse quesito, a que mais se destaca é a Methontology por ser praticamente completa em relação a um ciclo de desenvolvimento, não propondo apenas a fase de pré-desenvolvimento. Diante disso, a Methontology foi a que mais se destacou em propostas que poderiam vir a ser úteis à fase final desta pesquisa, conforme pode ser visualizado na Tabela 24.
- Em relação a detalhes das atividades e dos procedimentos para condução das mesmas, algumas metodologias e métodos mostram-se superficiais na elucidação dos passos para construção de ontologias. É o caso dos métodos Cyc, Kactus e Sensus, os quais parecem considerar que o *ontologista* já domina o assunto sobre construção de ontologias e não necessita de detalhes acerca de atividades e procedimentos envolvidos. Já a Methontology se destaca por fornecer, na maioria das vezes, detalhes de como proceder na condução de uma dada atividade.
- Algumas abordagens dão mais ênfase em atividades de desenvolvimento, especialmente na atividade de implementação da ontologia (método Cyc e método 101), desconsiderando aspectos importantes relacionados a gerenciamento do projeto, a estudo de viabilidade, a manutenção e a avaliação de ontologias.

- O manual da BITI apresentou um ciclo de vida praticamente completo segundo a norma IEEE-1074 (1997), conforme pode ser visto na Tabela 23. Apesar de as recomendações da norma ANSI/NISO Z39.19-2005 terem sido classificadas nas categorias de análise, tal norma não objetiva propor um ciclo de atividades para a construção de vocabulários controlados, e sim recomendar uma série de tratamentos em relação ao termo frente a atividades que se enquadram no processo de construção. Em ambos os casos, as fases de implementação e integração não foram consideradas por não pertencerem ao propósito dos vocabulários controlados. Desse modo, pode-se conferir a maturidade da metodologia exposta no manual da BITI frente a um modelo de ciclo de vida.

- Finalmente, é válido ressaltar que a metodologia proposta no manual da BITI para construção de tesauros se destacou perante as metodologias e aos métodos para construção de ontologias no quesito princípios teóricos e metodológicos para identificação, definição e organização de conceitos e de termos. Apesar de essas últimas indicarem métodos como *middle-out* para identificar conceitos, apresentarem recursos para se definir conceitos e organizá-los em uma taxonomia, não deixam explícitos os princípios teóricos que governam os seus construtos. Já no caso da construção de vocabulários controlados (especificamente de tesauros) é explícita a inserção de princípios da teoria da classificação, da teoria da terminologia e da teoria do conceito na especificação de seus elementos. Tais princípios seriam pertinentes em uma proposta metodológica para construção de ontologias, visto que ambos os instrumentos, ontologias e vocabulários controlados, representam relacionamentos semânticos e conceituais.

Fases do ciclo de vida			Cyc	Gruninger e Fox	Uschold e King	Kactus	Methontology	Sensus	Método 101	Z39.19-2005	Manual BITI
Gerenciamento do Projeto			Ausente	Ausente	Ausente	Ausente	Proposto	Ausente	Ausente	Proposto	Descrito em detalhe
Processos orientados	**Pré-desenvolvimento**		Ausente	Proposto	Ausente	Ausente	Ausente	Ausente	Ausente	Proposto	Ausente
	Desenvolvimento	**Especificação de requisitos**	Proposto	Descrito em detalhe	Proposto	Proposto	Descrito em detalhe	Ausente	Descrito em detalhe	Ausente	Proposto
		Modelagem conceitual	Ausente	Descrito em detalhe	Descrito em detalhe	Proposto	Descrito em detalhe	Proposto	Descrito em detalhe	Proposto	Descrito em detalhe
		Formalização	Ausente	Descrito em detalhe	Ausente	Ausente	Proposto	Proposto	Descrito em detalhe	Descrito em detalhe	Descrito em detalhe
		Implementação	Proposto	Proposto	Proposto	Ausente	Proposto	Proposto	Proposto	Ausente	Ausente
	Pós-desenvolvimento	**Manutenção**	Ausente	Ausente	Ausente	Ausente	Proposto	Ausente	Ausente	Descrito em detalhe	Proposto
Processos integrais	**Integração**		Proposto	Proposto	Proposto	Proposto	Proposto	Proposto	Proposto	Ausente	Ausente
	Avaliação		Ausente	Descrito em detalhe	Proposto	Ausente	Descrito em detalhe	Ausente	Ausente	Descrito em detalhe	Proposto
	Documentação		Ausente	Ausente	Proposto	Ausente	Descrito em detalhe	Ausente	Proposto	Proposto	Proposto

Tabela 23 – Tabela sinóptica da análise comparativa dos objetos investigados

Fases do ciclo de vida			Cyc	Gruninger e Fox	Uschold e King	Kactus	Methontology	Sensus	Método 101	Z39.19-2005	Manual BITI
Gerenciamento do Projeto			Ausente	Ausente	Ausente	Ausente	Processo de ciclo de vida baseado na evolução de protótipos. Escalonamento de atividades.	Ausente	Ausente	Ausente	Período de atualização; grupo para atuar nas manutenções.
Processos orientados	**Pré-desenvolvimento**		Ausente	Cenários de motivação	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente
	Desenvolvimento	**Especificação de requisitos**	Ausente	Questões de competência informal	Determinar o grau de formalismo	Ausente	Métodos e técnicas de Aquisição de conhecimento	Ausente	Questões de competência informal	Ausente	Delimitação da área; definição do público-alvo.
		Modelagem conceitual	Ausente	Ausente	Tratamento de termos sinônimos e ambíguos.	Reuso de conhecimento de ontologias de alto nível	Métodos para a construção das representações intermediárias	Reuso de conhecimento de ontologias de alto nível	Conjunto de princípios para construção de hierarquias de classes	Recomendações para o tratamento da terminologia	Análise do domínio segundo as teorias do conceito e da classificação
		Formalização	Ausente	Questões de competência. Especificação de axiomas	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente
		Implementação	Ausente	Ausente	Critérios para escolha da linguagem	Ausente	Critérios para escolha de ambientes de desenvolvimento	Ausente	Uso interativo da linguagem através do *Protégé*	Ausente	Ausente
	Pós-desenvolvimento	**Manutenção**	Ausente	Ausente	Ausente	Ausente	Verificações nas representações intermediárias	Ausente	Ausente	Ausente	Ausente
Processos integrais	**Integração**		Ausente	Ausente	Ausente	Ausente	Documento de integração com meta-ontologias	Ausente	Ausente	Ausente	Ausente
	Avaliação		Ausente	Questões de competência	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente
	Documentação		Ausente	Ausente	Ausente	Ausente	Documentação em cada fase; representações intermediárias.	Ausente	Ausente	Ausente	Ausente

Tabela 24 – Mapeamento de características que poderiam contribuir à fase final da pesquisa

6 - Proposta Metodológica para Construção de Ontologias

O presente capítulo apresenta, finalmente, a etapa de desenvolvimento da proposta metodológica para construção de ontologias, a fim de efetivar o objetivo geral da pesquisa.

No capítulo 4 foram executadas as devidas coletas e análises sobre o conteúdo do material empírico da pesquisa, incluindo o método Cyc, a metodologia de Gruninger e Fox, o método de Uschold e King, o método Kactus, a metodologia Methontology, o método Sensus, o método 101, a metodologia do manual da BITI para construção de tesauros e a norma da ANSI/NISO Z39.19-2005 para construção de vocabulários controlados.

A partir da apresentação dos resultados do capítulo 4, expôs-se no capítulo 5 a análise comparativa sobre os objetos investigados e foram tecidas considerações acerca das comparações. Tais considerações foram essenciais para a execução da presente etapa de pesquisa, construída sobre o arcabouço conceitual anteriormente delineado e sobre as análises comparativas realizadas.

Este capítulo encontra-se organizado da seguinte forma: a seção 6.1 tece as considerações iniciais sobre a proposta metodológica no sentido de explicitar os pontos de partida no desenvolvimento da proposta; a seção 6.2 discorre sobre as fases constituintes da proposta metodológica; e, finalmente, a seção 6.3 faz as considerações interdisciplinares da proposta sintetizando-as em um panorama interdisciplinar de princípios teóricos e metodológicos das Ciências da Informação e da Computação.

6.1 Considerações iniciais sobre a proposta metodológica

Antes de iniciar o desenvolvimento da proposta metodológica para construção de ontologias, achou-se pertinente apresentar os <u>requisitos</u> considerados fundamentais para a consecução de tal proposta.

O primeiro requisito está direcionado à estratégia de construção das ontologias, a qual se embasa em um <u>modelo de ciclo de vida</u> fundamentado na norma internacional IEEE-1074 (1997).

O segundo requisito está voltado à estratégia de construção de <u>ontologias de domínio</u>, as quais precisariam de uma conceitualização formada por conceitos de alto nível e conceitos comuns ao domínio.

O terceiro requisito vislumbra um ciclo de vida que contemple o momento em que as atividades devem ser conduzidas, ou seja, a identificação de um conjunto de fases relevantes para o processo de construção; <u>quais</u> atividades são executadas em cada uma destas fases; e <u>como</u> estas

fases estão relacionadas, incluindo a relação de precedência e os formatos dos produtos que caracterizariam o final de cada uma delas, e o início da fase seguinte.

O quarto e último requisito demanda a explicitação dos princípios teóricos e metodológicos de cada campo do conhecimento envolvido na proposta.

Postas as considerações, as próximas seções se encarregam de apresentar a proposta.

6.2 As fases da proposta

Conforme explicitado no primeiro requisito da proposta, as fases da metodologia para construção de ontologias estão associadas a um modelo de ciclo de vida advindo da norma IEEE-1074 (1997), que descreve o processo de desenvolvimento de maneira estruturada e metódica. A proposta metodológica para construção de ontologias inclui fases que vão desde o conhecimento da situação atual do domínio até a manutenção da ontologia. A Figura 70 apresenta as fases da proposta que seguem a ordem explicitada na norma em questão: *gerenciamento do projeto; pré-desenvolvimento; processos de desenvolvimento; pós-desenvolvimento e processos integrais*.

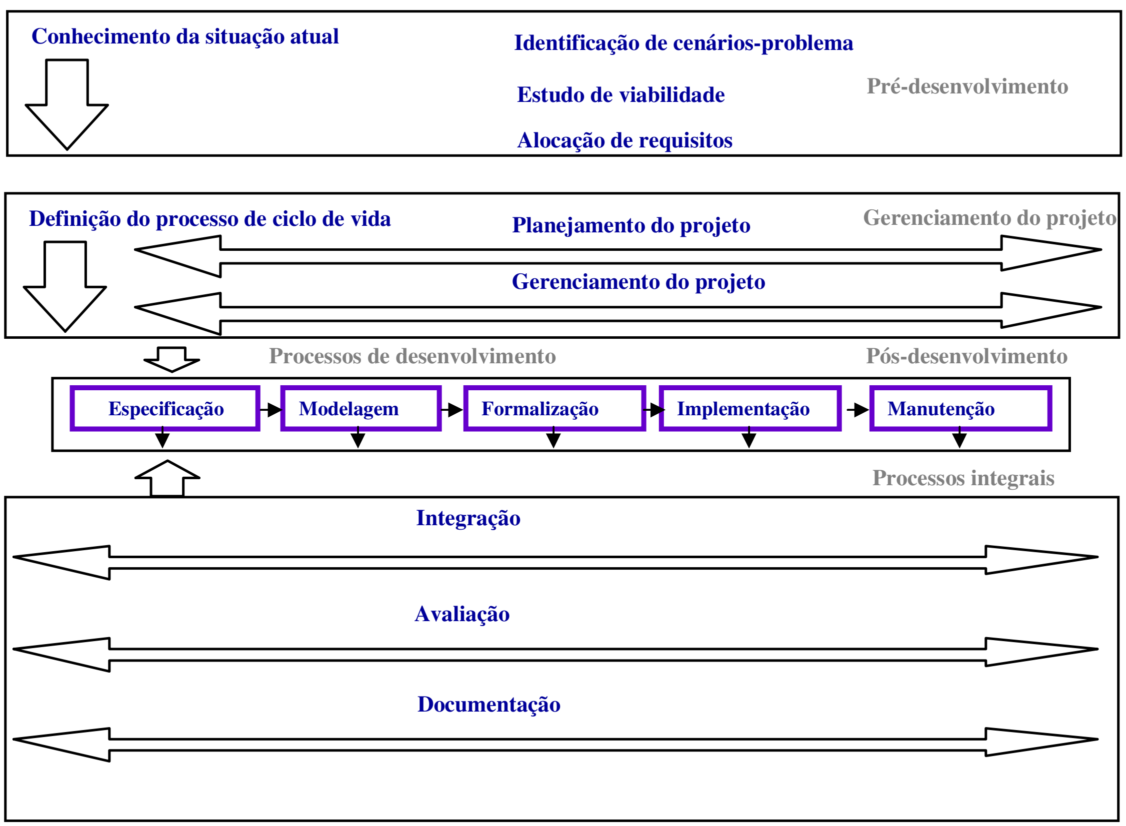


Figura 70 – Fases da proposta metodológica para construção de ontologias

Em cada fase da proposta é apresentado o fluxo de informação (entrada e saída de produtos) entre as atividades alocadas, sendo que tal fluxo advém da norma IEEE-1074 (1997). Informação ou produto de entrada consiste em uma lista de produtos requeridos na fase, que serão transformados

através de atividades realizadas na fase; e informação ou produto de saída consiste em uma lista de produtos resultantes de uma fase, que servirão como entrada em outra(s) fase(s).

As fases envolvendo a proposta metodológica foram estruturadas de acordo com as contribuições advindas do mapeamento de características de cada um dos objetos estudados, como sumarizado na Tabela 24 do capítulo 5. Nestas características puderam-se notar as grandes contribuições das teorias advindas da Ciência da Informação – como a teoria da classificação facetada e a teoria do conceito – e da Ciência da Computação – advindas de áreas como a Engenharia de Software, a Engenharia do Conhecimento e a Engenharia ontológica. Tais teorias serão destacadas na elucidação das fases da proposta metodológica para construção de ontologias, apresentadas nas próximas seções.

Destaca-se que, para efeitos didáticos, na proposta de elaboração de modelos para registro de informações, incluindo as representações intermediárias para a construção da conceitualização da ontologia, foram apresentados exemplos acerca de um domínio de viagem extraídos do trabalho de Fernandez, Gomez-Perez e Corcho (2004).

6.2.1 Fase de Gerenciamento do Projeto

A fase de gerenciamento do projeto demanda atividades relacionadas ao início, ao planejamento e a gestão do projeto da ontologia em todo o seu ciclo de vida. O produto requerido na fase pode ser conferido na Tabela 25.

Informação de entrada	Origem	
	Fase	Atividade
Requisitos para desenvolvimento	Pré-desenvolvimento	Alocação de requisitos
Recomendação de melhoria no processo de desenvolvimento da ontologia	Manutenção	Manutenção nas fases
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 25 – Produtos requeridos na fase de gerenciamento do projeto

A atividade de início do projeto consiste na criação do processo de ciclo de vida da ontologia e é descrita na seção 6.2.1.1. A atividade de planejamento do projeto inclui um escalonamento que identifica as tarefas a serem executadas em cada fase, a maneira como são organizadas e os recursos necessários para a realização de cada uma delas. Tal atividade é descrita na seção 6.2.1.2. E, finalmente, a atividade de gestão do projeto garante que as tarefas escalonadas

sejam completadas de modo adequado e que seus recursos sejam utilizados conforme o planejado. Tal atividade é descrita na seção 6.2.1.3.

6.2.1.1 Criação do processo de ciclo de vida

A atividade de criação do processo de ciclo de vida envolve a decisão de qual processo é o mais adequado ao desenvolvimento de ontologias. Conforme discutido na seção 4.5 (análise da metodologia Methontology), o processo de construção de ontologias é similar aos modelos de desenvolvimento em cascata, incremental e evolutivo (PRESSMAN, 2002). Entretanto, pôde-se perceber que o processo de ciclo de vida baseado na evolução de protótipos é o mais adequado à construção de ontologias (FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997), no sentido de permitir a modificação, a inclusão e a remoção de definições conceituais na ontologia a qualquer momento a partir da necessidade de desenvolvimento de uma nova versão (protótipo).

A proposta metodológica, então, recomenda o processo de ciclo de vida baseado na evolução de protótipos, apresentado na Figura 71, o qual demanda um planejamento e gerenciamento para ser bem sucedido.

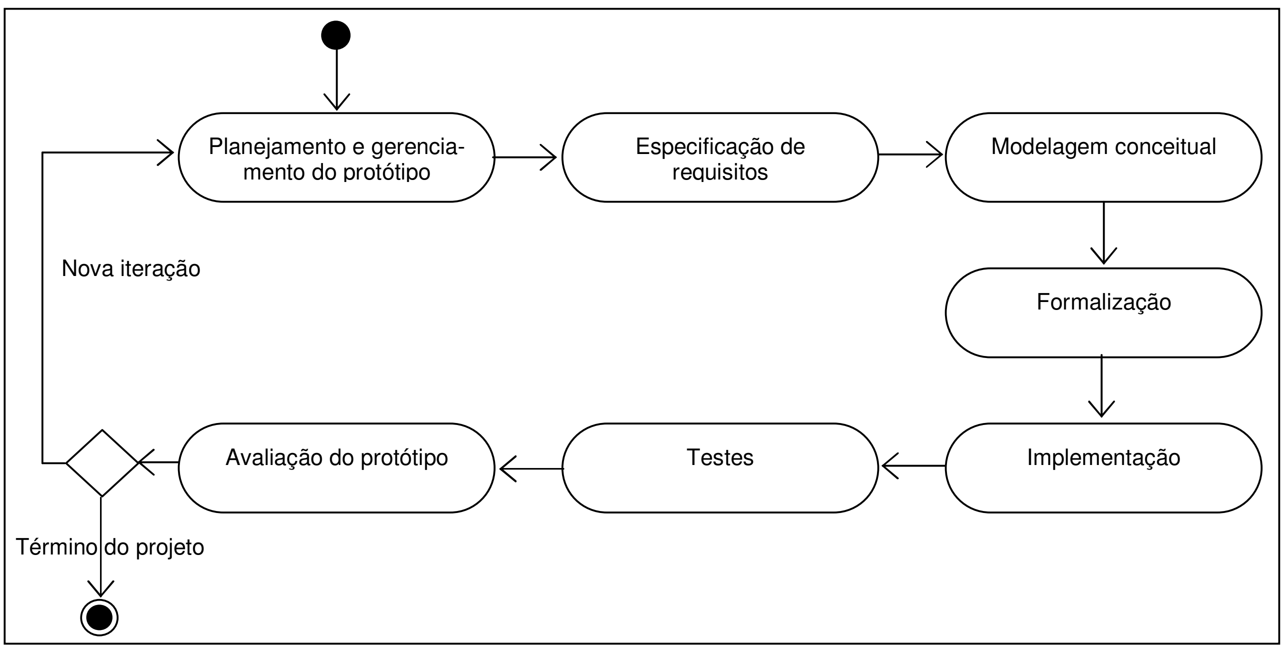


Figura 71 – Processo de ciclo de vida baseado na evolução de protótipos

6.2.1.2 Planejamento do protótipo

A atividade de planejamento do protótipo consiste em determinar <u>quais</u> atividades, juntamente com suas tarefas, deverão ser realizadas em cada fase do projeto; <u>quem</u> deverá realizá-las; com <u>quais</u> recursos; <u>quando</u> devem ser iniciadas e finalizadas; e <u>quanto</u> elas deverão custar. Tudo isto requer a elaboração de estimativas em relação ao número e à dimensão dos produtos

produzidos em cada fase, do número de pessoas necessárias, dos prazos e dos custos. Sem um planejamento adequado, os *ontologistas* envolvidos no projeto não saberão o que fazer e não terão datas para iniciar ou terminar o trabalho. Sem estimativas, o responsável pelo planejamento não terá como dimensionar o tamanho e a quantidade de produtos a serem produzidos e o esforço necessário para produzi-los. Por fim, sem um orçamento, não se terá noção de quanto a ontologia irá custar e se haverá recursos para o seu desenvolvimento. Diante disso, a atividade de planejamento demanda as seguintes tarefas (PRESSMAN, 2002; SOMMERVILLE, 2007):

- Realizar estimativas.
- Elaborar o escalonamento de atividades em cada fase, incluindo tarefas, pessoas alocadas, outros recursos necessários e orçamento previsto. Esta tarefa é conduzida de acordo com o produto fornecido pela fase pré-desenvolvimento (discutida na próxima seção), em que uma análise de ambiente é feita no sentido de alocar os requisitos que serão utilizados no processo de desenvolvimento da ontologia.
- Elaborar cronograma.
- Elaborar orçamento.

A Tabela 26 apresenta um fragmento de um modelo proposto para planejar o escalonamento de atividades envolvidas no ciclo de vida da ontologia. Em cada fase são identificadas as atividades envolvidas, a previsão de orçamento, as tarefas em ordem cronológica de execução, a data inicial e final para a realização, as pessoas alocadas e outros recursos necessários.

Fase 1						
	Atividade 1				**Orçamento previsto**	
		Tarefa 1				
			DT Inicial	DT final	Alocação pessoal	Outros recursos
		Tarefa 2				
			DT Inicial	DT final	Alocação pessoal	Outros recursos
	Atividade 2				**Orçamento previsto**	
		Tarefa 1				
			DT Inicial	DT final	Alocação pessoal	Outros recursos
Fase 2						
	Atividade 1				**Orçamento previsto**	
		Tarefa 1				
			DT Inicial	DT final	Alocação pessoal	Outros recursos
		Tarefa 2				
			DT Inicial	DT final	Alocação pessoal	Outros recursos
		Tarefa 3				
			DT Inicial	DT final	Alocação pessoal	Outros recursos

Tabela 26 – Modelo de escalonamento de atividades para as fases do ciclo de vida da ontologia

A atividade de planejamento do protótipo resulta em um plano de desenvolvimento que inclui o modelo de escalonamento apresentado anteriormente, além de informações sobre o planejamento da documentação, da avaliação e da manutenção em cada fase do projeto.

6.2.1.3 Gerenciamento do protótipo

A atividade de gerenciamento do protótipo objetiva fazer cumprir o que foi planejado. O papel do gerente de protótipos da ontologia é coordenar a equipe, controlar a produção da documentação em todas as fases, controlar as versões dos protótipos da ontologia, fazer cumprir prazos e custos previstos e analisar métricas de produção.

O gerenciamento do protótipo está diretamente ligado à garantia da qualidade do processo e da ontologia. O uso de métricas de qualidade (PRESSMAN, 2002, p.90), tanto da ontologia como do processo, é fundamental para o gerenciamento. Com base em métricas, o gerente tem condições de avaliar se o planejamento do protótipo está sendo cumprido ou não. Neste caso, as métricas podem apontar as causas dos problemas e permitir as revisões no planejamento. Exemplos de métricas da ontologia são: tamanho da estrutura hierárquica; número de representações intermediárias na atividade de conceitualização; número de inconsistências nas representações formais; entre outras. Exemplos de métricas[139] de processo são: o esforço; a produtividade; entre outras.

A atividade de gerenciamento do protótipo resulta em um plano de gerenciamento que inclui todos os métodos e técnicas utilizadas no processo de gerenciamento em cada fase do projeto.

A Tabela 27 apresenta os produtos resultantes da fase de gerenciamento do projeto e suas fases e atividades requerentes. Ressalta-se que os planos são direcionados para todas as fases a fim de serem seguidos, entretanto, no caso específico da fase de avaliação, tais planos também são submetidos a uma avaliação acerca de seus conteúdos.

Informação de Saída	Destino	
	Fase	Atividade
Plano de desenvolvimento	Todas	Todas
Plano de gerenciamento	Todas	Todas

Tabela 27 – Produtos resultantes da fase de gerenciamento do projeto

[139] Informações adicionais sobre métodos direcionados a métricas de processo e projeto de *software* podem ser consultadas em Pressman (2002, p.75).

6.2.2 Fase Pré-desenvolvimento

A fase pré-desenvolvimento da ontologia envolve profissionais (*ontologista* e especialista do domínio) que buscam o conhecimento da situação atual do domínio e a identificação de problemas (exploração de idéias ou conceitos pela IEEE-1074) a fim de elaborar propostas de solução através de ontologias. O uso de técnicas de análise de sistemas advindas da área de Engenharia de Software pode ser vantajoso, no sentido de levantar os problemas e apresentar alternativas de soluções. Tais alternativas são acompanhadas de um estudo da viabilidade de cada uma delas, incluindo análise custo-benefício para se decidir qual solução será a escolhida.

A Tabela 28 apresenta os produtos necessários para a realização e a concretização das atividades envolvidas em tal fase.

Informação de entrada	Origem	
	Fase	Atividade
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 28 - Produtos requeridos na fase pré-desenvolvimento

A técnica proposta para a identificação de problemas em um domínio é a análise de cenários (BENNER et al., 1993), conforme discutido na seção 2.3.3.5.2. Tal técnica favorece o levantamento de informações, a identificação de problemas e a antecipação das soluções. Para tal, antes de descrever os cenários, *o ontologista* deve entrevistar o especialista para entender os problemas e os requisitos iniciais. Desse modo, a entrevista permite que o especialista descreva as suas tarefas, os problemas associados a cada uma delas e as possíveis soluções.

A proposta metodológica recomenda mapear e documentar os *cenários-problema*, incluindo informações sobre o período de aquisição do conhecimento, os participantes da atividade de análise (*ontologista* e especialista), a descrição do cenário, o impacto causado pelo problema, a solução proposta e observação.

A Tabela 29 apresenta a proposta de um modelo para a documentação de um *cenário-problema*, e ilustra um exemplo de vendas de pacote de viagem em uma agência de turismo.

Cenário: 001	Breve descrição: Vendas de pacote de viagem em uma agência de viagem
Período: xx/xx/xxxx a xx/xx/xxxx	
Ontologista: xxxxxxxxxxxxxxxxxxxxxxxxxxxx	
Especialista: xxxxxxxxxxxxxxxxxxxxxxxxxxxx	
Descrição do cenário: Em uma agência de viagem, um cliente solicita um pacote de viagem, desejando que a agência o indique opções de lugares mais apropriados para a viagem.	
Problema identificado: Descobrir os destinos mais apropriados para uma viagem.	
Impacto do problema: Indicação indevida de lugares-destino.	
Proposta de solução: Construir uma ontologia para representar o conhecimento sobre viajantes, viagens e destinos.	
Observação:	

Tabela 29 – Modelo de especificação de cenários-problema

Para a condução do estudo de viabilidade (PRESSMAN, 2002; SOMMERVILLE, 2007) a proposta metodológica recomenda verificar: i) se é possível construir a ontologia com o orçamento e tecnologia disponível; ii) se a ontologia pode ser integrada com outras ontologias; e iii) se a ontologia será desenvolvida do início ou através da extensão de uma já existente. Além disso, questões podem ser elaboradas no sentido de auxiliar o estudo de viabilidade, tais como:

- O que aconteceria se a ontologia não fosse construída?
- Quais são os problemas com o conhecimento atual?
- Como a ontologia proposta irá ajudar?
- Quais seriam os problemas de integração?
- É necessária nova tecnologia? Quais habilidades são requeridas?
- Quais facilidades devem suportar a ontologia proposta?

Na atividade de análise, os profissionais envolvidos também buscam identificar e alocar requisitos que serão utilizados no processo de desenvolvimento da ontologia. Os requisitos são fundamentais para que o gerente de projeto possa elaborar um plano de desenvolvimento, indicando em detalhe os recursos necessários (*hardware*, *software*, pessoal, procedimentos, informação, documentação, etc), bem como as estimativas de prazos e custos (cronograma e orçamento).

Recomenda-se atualizar o modelo de especificação de *cenários-problema* sempre que um novo problema ocorrer no domínio e necessitar de solução por meio de ontologias.

A Tabela 30 mostra os produtos resultantes da fase pré-desenvolvimento e as fases e atividades dependentes dos mesmos.

Informação de saída	Destino	
	Fase	Atividade
Requisitos para desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Cenários-problema	Especificação de requisitos	Identificação de questões de competência
	Avaliação	Avaliação de resultados

Tabela 30 – Produtos resultantes da fase pré-desenvolvimento

6.2.3 Fase de Especificação de Requisitos

A fase de especificação de requisitos é o ponto de partida para a construção da ontologia. O conjunto de *cenários-problema* determinado na fase pré-desenvolvimento, conforme mostra a Tabela 31, possibilita ao *ontologista* identificar um conjunto de questões de competência (GRUNINGER e FOX, 1995), de forma que os problemas levantados possam ser convertidos em questões que a ontologia deverá ser capaz de resolver.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Cenários-problema	Pré-desenvolvimento	Análise da situação atual do domínio
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação	Documentação	Produção e distribuição de documentos

Tabela 31 – Produtos requeridos na fase de especificação de requisitos

Na atividade de elaboração das questões de competência, o *ontologista*, em conjunto com especialistas do domínio, deve analisar os problemas e as soluções propostas (registrados no modelo de especificação de *cenários-problema*) e produzir questões e respostas em linguagem natural. A técnica de análise de cenários também pode ser empregada na especificação das questões de competência e de suas respostas, pois auxilia os envolvidos no levantamento de informações perante situações reais ocorridas em um domínio. Finalmente, cada questão com sua respectiva resposta deve ser registrada no modelo de especificação (conseqüentemente modificado), conforme é mostrado de forma resumida na Tabela 32. Tais questões e respostas servirão de base para a extração e definição do conhecimento necessário à fase de modelagem conceitual (discutida na próxima seção), em que a conceitualização da ontologia é concebida.

Cenário: 001	Breve descrição: Vendas de pacote de viagem em uma agência de viagem
Problema identificado: Descobrir os destinos mais apropriados para uma viagem.	
Questão de competência: Considerando as preferências de um viajante (viagem por cultura, para as montanhas, para a praia, etc) e algumas restrições (econômica ou sobre a própria viagem), *quais destinos são mais apropriados*?	
Resposta à questão: Conhecer as preferências de um viajante, as restrições sobre a viagem e os destinos disponíveis.	

Tabela 32 – Questão de competência e sua resposta para um *cenário-problema*

As questões de competência também são usadas para avaliar o compromisso ontológico estabelecido na definição da ontologia, de forma a verificar se a mesma satisfaz os requisitos especificados. A avaliação da ontologia será discutida na fase de avaliação da seção 6.2.9.

Diante às questões que a ontologia precisará responder torna-se possível determinar o domínio em que a ontologia irá cobrir e o propósito de construção, incluindo o uso pretendido da mesma. No exemplo fornecido na fase pré-desenvolvimento, o propósito de se construir uma ontologia seria em proporcionar uma base de conhecimento consensual sobre o domínio de viagem, que pode ser usada em agências de viagem. Além disso, o *ontologista* e os especialistas precisam definir qual é o grau de formalismo pretendido sobre a ontologia, podendo ser de informal, utilizando linguagem natural, a formal, usando formalismos lógicos (USCHOLD e GRUNINGER, 1996). Tais definições resultam em um documento preliminar denominado Escopo da ontologia, o qual registra informações pertinentes ao conhecimento sobre a ontologia. Tal documento vai sendo refinado na medida em que as fases vão sendo concretizadas, conforme pode ser visualizado na Figura 72, em que um modelo de escopo é proposto.

Escopo da Ontologia
Domínio: Viagem
Data: xx de xxxx de xxxx
Conceituado por: preenchido na fase de modelagem
Formalizado por: preenchido na fase de formalização
Implementado por: preenchido na fase de implementação
Propósito:
Proporcionar uma base de conhecimento consensual sobre o domínio de viagem, que pode ser usada em agências de viagem.
Nível de formalidade: formal
Lista de termos:
Preenchido na fase de modelagem
Fontes de conhecimento:
Preenchido na fase de modelagem

Figura 72 – Modelo de escopo da ontologia

É válido ressaltar que esta proposta recomenda um ciclo de vida baseado em evolução de protótipos, conforme explanado na fase de gerenciamento do projeto. Sendo assim, o modelo de especificação de *cenários-problema* pode resultar, em princípio, num documento parcialmente completo.

A Tabela 33 mostra os produtos gerados na fase de especificação de requisitos e as fases e atividades que necessitam de tais produtos.

Informação de saída	Destino	
	Fase	Atividade
Questões de competência	Modelagem conceitual	Identificação de termos
		Definição de termos
	Formalização	Formalização de questões
Respostas às questões de competência	Modelagem conceitual	Identificação de termos
		Definição de termos
Escopo parcial da ontologia	Modelagem conceitual	Identificação de termos
	Formalização	Formalização
	Implementação	Implementação da ontologia
	Avaliação	Avaliação de resultados

Tabela 33 - Produtos resultantes da fase de especificação de requisitos

6.2.4 Fase de Modelagem Conceitual

O propósito da fase de modelagem conceitual é conceber a conceitualização da ontologia. Para tal, reporta-se aos componentes de tal instrumento (GRUBER, 1993a; GOMEZ-PEREZ, FERNANDEZ e VICENTE, 1996), de forma que os mesmos possam ser organizados e representados como uma estrutura de representação do conhecimento. A saber: a) **classes conceituais**, que organizam os conceitos de um domínio em uma taxonomia; b) **atributos de classes**, que são propriedades relevantes do conceito; c) **instâncias**, que são utilizadas para representar objetos específicos de um conceito; d) **atributos de instâncias**, que são propriedades relevantes que descrevem as instâncias de um conceito; e) **relações entre classes**, que representam o tipo de interação entre os conceitos de um domínio; f) **constantes**, que têm sempre o mesmo valor e, geralmente, são usadas em regras ou fórmulas para inferir conhecimento na ontologia; g) **termos**, que designam os conceitos de um domínio; h) **axiomas formais**, que restringem a interpretação e o uso dos conceitos envolvidos na ontologia; e i) **regras**, que determinam condições ao domínio, além de inferirem valores para atributos.

Tendo em vista que a taxonomia é a base das ontologias, e que tal estrutura é constituída de termos representativos dos conceitos do domínio, torna-se necessário fazer o tratamento de tais termos e conceitos envolvidos e, logo em seguida, conduzir a organização dos mesmos na estrutura taxonômica. Na atividade de criação das classes conceituais, o primeiro passo é buscar os termos que representam o assunto da ontologia (definido na fase de especificação de requisitos). Uma vez

obtida uma lista de termos, estes são definidos e agrupados em categorias de acordo com sua natureza, formando-se assim as classes conceituais. Por fim, esses termos são relacionados entre si, resultando em uma sistematização de todo o conhecimento que está sendo construído.

Nesta fase da proposta metodológica para construção de ontologias, indica-se a utilização do método de análise facetada na elaboração de classificações, com princípios da teoria da classificação facetada (RANGANATHAN, 1967), e da análise de domínio na identificação de termos e definições de conceitos, relações e atributos, utilizando princípios da teoria do conceito (DALHBERG, 1978) e da teoria da terminologia (WÜESTER, 1981[140] apud CAMPOS, 2001). Pretende-se usar ainda, como apoio ao processo de modelagem, métodos e técnicas de algumas metodologias e métodos para construção de ontologias (mapeados na tabela 24 do capítulo 5), destacando as representações intermediárias propostas na Methontology (GOMEZ-PEREZ, FERNANDEZ e VICENTE, 1996; FERNANDEZ, GOMEZ-PEREZ e CORCHO, 2004) na especificação de conceitos, relações, atributos, instâncias, dentre outros aspectos relevantes ao processo de construção da conceitualização. Além disso, serão observados alguns princípios para definição de classes e hierarquias de classes no processo de concepção de taxonomias, advindos do método 101 (NOY e McGUINNESS, 2001).

Campos (2004) discorre que o método de "pensar" sobre o domínio na teoria do conceito pode ser considerado híbrido na agregação do método dedutivo e do método indutivo. No primeiro formam-se categorias para representar contextos, e no segundo analisam-se os conceitos de um contexto no propósito de ordená-los no interior de tais categorias. O método de Dahlberg preocupa-se primariamente com o entendimento do conceito, e em seguida com a classificação deste em um domínio de conhecimento.

A análise facetada propicia ao *ontologista* refletir sobre uma dada realidade (um domínio), de forma abstrata, independentemente do arranjo particular dos objetos no domínio. Isto seria, portanto, o primeiro passo na concepção da conceitualização da ontologia, que será representada por uma estrutura taxonômica. No processo de análise facetada é adotada a metodologia analítico-sintética para organização do conhecimento do domínio. Para tal, os conceitos identificados e definidos (processo de análise) são classificados em categorias (processo de síntese), em que cada categoria representa uma característica. Conforme assinala Campos (2004, p.27), o agrupamento de objetos através de categorias possibilita diminuir erros lógicos no estabelecimento das ligações entre os conceitos, pois determina a natureza do objeto. Gomes et al. (1990, p.37) ainda acrescenta que a vantagem do método de faceta é possibilitar a formação de categorias mutuamente exclusivas, isto é, quando se tem um termo novo, sabe-se prontamente em que categoria incluí-lo.

---

[140] WUESTER, E. L. L'étude scientifique qénérale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, L'informatique et les sciences des chose. In: RONDEAU, G. ; FELBER, E. (Org.). Textes choisis de Terminologie. Québec: GIRSTERM, 1981. p. 57-114.

Para a formação das categorias, a proposta metodológica deste estudo considera três frentes: i) as categorias fundamentais PMEST propostas por Ranganathan em sua teoria da classificação facetada; ii) as categorias fundamentais estendidas posteriormente pelo *Classification Research Group* – CRG para construção de tesauros; e iii) as categorias e as subcategorias formais de Dalhberg em sua teoria do conceito.

As cinco categorias fundamentais Personalidade, Matéria, Energia, Espaço e Tempo permitem a divisão de qualquer assunto de qualquer área do conhecimento utilizando, para tal, um método dedutivo, no qual se considera primeiro o domínio ou contexto para depois fazer o tratamento dos elementos que compõem tal domínio. Desse modo, o método possui mecanismos de representação para trabalhar com metaníveis conceituais – as categorias predefinidas e, a partir delas, os conceitos são ordenados para formar classes de conceitos.

As dez categorias fundamentais estendidas pelo CRG são de natureza flexível por serem moldadas a partir de assuntos específicos, a saber: i) coisas, substâncias, entidades; ii) suas partes; iii) sistemas de coisas; iv) atributos de coisas; v) objeto da ação; vi) relações entre coisas; vii) operações sobre coisas; viii) propriedades de atributos, relações e operações; ix) lugar, condição; e x) tempo. Os tesauros baseados em facetas geralmente utilizam as categorias propostas pelo CRG, conforme foi visto nas seções 2.2.3.3.3 e 4.8.1.3 deste estudo.

As categorias e as subcategorias formais propostas por Dalhberg são em número superior às fundamentais de Ranganathan. A proposta de Dahlberg recomenda quatro categorias e doze subcategorias, a saber: Entidades (Princípios, Objetos imateriais e Objetos materiais), Propriedades (Quantidades, Qualidades e Relações), Atividades (Operações, Estados e Processos) e Dimensões (Tempo, Posição e Espaço).

As categorias das três propostas se manifestam através de facetas, que agrupam conceitos sob a luz de um princípio de divisão comum, mutuamente exclusivo (RANGANATHAN, 1967; BARBOSA, 1972; DALHBERG, 1976; VICKERY, 1980, GUINCHAT e MENOU, 1994). Uma observação importante feita por Campos, Gomes e Motta (2004) é que nem todos os domínios possuem manifestações (facetas) para todas as categorias.

Tendo em vista que o propósito das ontologias de domínio, foco desta proposta metodológica, é descrever um vocabulário relacionado a um domínio genérico (GUARINO, 1998), e que a organização de seus elementos se dá a partir de conceitos de alto nível e de conceitos comuns ao domínio, julgou-se pertinente recomendar categorias ou facetas que contemplassem tais conceitos, os quais poderiam ser utilizados independentemente de suas particularidades. Desse modo, dentre as três propostas aventadas e anteriormente comentadas, a que mais se enquadrou em tal perspectiva foi a de Dahlberg (DALHBERG, 1978) com suas categorias e subcategorias abrangentes e definidas com clareza. Conforme discutido, o método da teoria do conceito considera

a abordagem dedutiva e indutiva no processo de análise sobre um domínio, indo mais ao encontro da realidade de construção de ontologias, na qual se utilizam de métodos *top-down* (baseados em processos dedutivos) para determinação de conceitos genéricos ou de alta abstração e *bottom-up* (baseados em processos indutivos) na determinação de conceitos de nível mais baixo. Tais métodos foram apresentados, na seção 4, em algumas metodologias e métodos para construção de ontologias.

Nessa proposta, as categorias manifestadas nas facetas servem de instrumento facilitador ao entendimento da natureza dos conceitos e das relações entre os conceitos, facilitando, dessa forma, a atividade de definição de termos. Na concepção da taxonomia, uma ontologia de alto nível é requerida como alternativa de reuso dos conceitos mais genéricos que irão compor a terminologia da ontologia de domínio. Desse modo, o esquema facetado pode auxiliar o cotejamento entre as categorias sintetizadas e as categorias da ontologia de alto nível, visto que as primeiras estão organizadas em metaníveis conceituais, aproximando-se, desse modo, dos conceitos da ontologia de alto nível. Além disso, o esquema facetado pode auxiliar o *ontologista* na identificação de termos sinônimos e ambíguos, termos genéricos e termos específicos dentro de cada faceta.

Tendo elucidado os princípios teóricos e metodológicos utilizados da fase de modelagem conceitual, as próximas seções se encarregam de apresentar as atividades necessárias ao tratamento e à organização de termos na estruturação do conhecimento sobre o domínio, isto é, na concepção da conceitualização. A Tabela 34 apresenta os produtos necessários à realização de tais atividades e a Figura 73 apresenta a seqüência de execução das mesmas.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Questões de competência	Especificação de requisitos	Elaboração das questões de competência
Respostas às questões	Especificação de requisitos	Elaboração das questões de competência
Escopo parcial da ontologia	Especificação de requisitos	Definição de escopo
Documento de integração	Integração	Integração com ontologias de alto nível
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 34 - Produtos requeridos na fase de modelagem conceitual

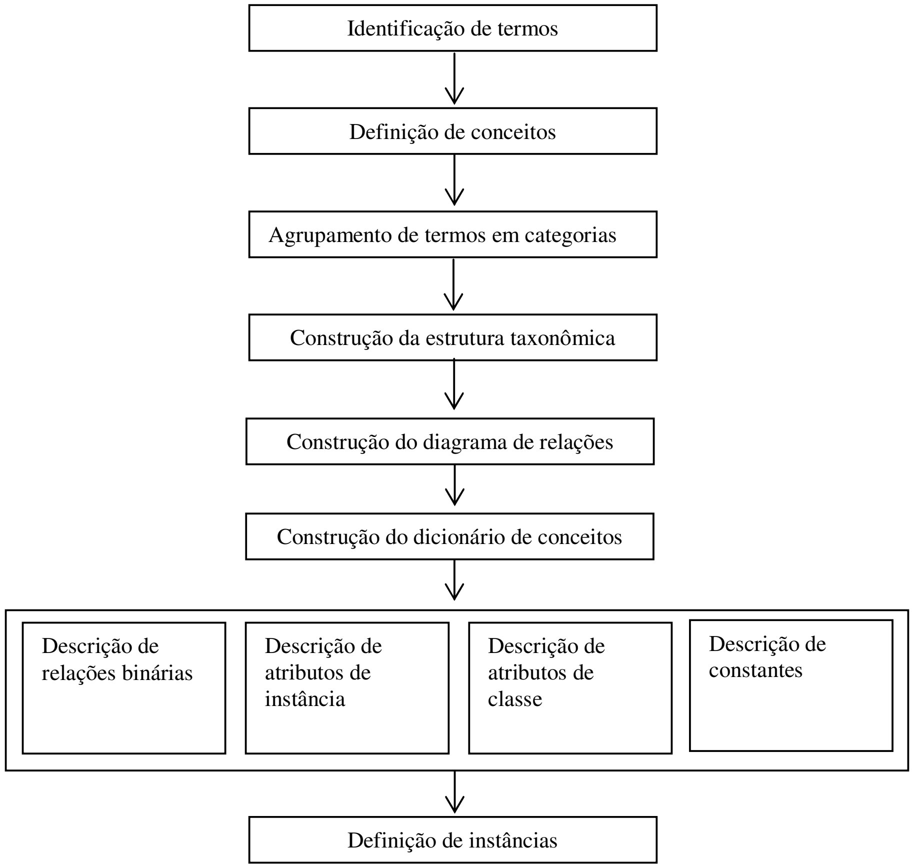


Figura 73 – Atividades para a concepção da conceitualização

6.2.4.1 Identificação de termos

Para o passo de identificação de termos representativos do assunto da ontologia utilizam-se princípios advindos da teoria geral da terminologia e da teoria do conceito, que auxiliam na escolha dos conceitos em um dado universo de conhecimento. Dahlberg (1978) em seu modelo para construção de conceitos (ver seção 2.2.3.2) propõe iniciar o processo pelo referente, o qual se encontra presente em um universo de itens, incluindo idéias, objetos, fatos, leis, propriedades, ações, dentre outros. Tais itens podem ser classificados em objetos que conseguem ser percebidos (ex: homem, vegetal e fruta); objetos considerados abstratos (ex: processo, atividade e tarefa); e objetos individuais ou instâncias de um conceito (ex: planetas terra, marte, júpiter do conceito sistema solar). O outro princípio, da contextualização, advindo da teoria geral da terminologia, determina a escolha dos conceitos de acordo com a área do conhecimento, que irá implicar no ponto de vista pelo qual o referente será observado no intuito de determinar as suas características.

No processo de análise das questões de competência e de suas respostas, registradas no modelo de especificação de *cenários-problema* da fase de requisitos, o *ontologista* deve extrair os termos potencialmente relevantes, que representam o conhecimento (conceitos, atributos e relações)

do domínio. Os termos identificados devem compor a lista de termos no documento de escopo da ontologia, que posteriormente deverá ser submetida a uma análise criteriosa de definições a partir de seu contexto (descrita na seção 6.2.4.2).

As possíveis fontes para ajudar na extração do conhecimento podem ser: especialistas do domínio, livros, manuais, figuras, tabelas, outras ontologias, tesauros, dentre outras. E os possíveis métodos e técnicas (advindos das áreas de Engenharia de Software, Engenharia ontológica, Biblioteconomia e Inteligência Artificial) utilizados na construção consensual do conhecimento podem ser: *Joint Application Design* – JAD, *brainstorming*, análise de documentos e mecanismos automatizados para extração de termos. O método JAD possibilita discussões em grupo ao invés de individual, possibilitando decisões consensuais acerca da terminologia do domínio. A sessão de *brainstorming* é uma técnica para especificar termos e frases relevantes para o domínio através de explanações de idéias entre especialistas (USCHOLD e GRUNINGER, 1996). Para a análise de documentos (incluindo a documentação produzida na fase de especificação de requisitos) poder-se-ia usar a técnica de análise de assunto (UNISIST, 1981; FUJITA, 2003), amplamente utilizada na Biblioteconomia e Ciência da Informação, a qual auxilia na análise e apreensão do conteúdo de documentos, possibilitando, desse modo, identificar e selecionar os conceitos que representam a essência de um documento. A utilização de instrumentos tecnológicos como a ferramenta *Domain Analysis and Reuse Environment* - DARE[141] (PRIETO-DÍAZ, 2003) podem ser úteis na extração de termos e frases em documentos.

Conforme discutido na fundamentação teórica e apresentado na análise das metodologias e métodos para construção de ontologias, a inspeção de meta-ontologias para o reaproveitamento de termos pode ser uma alternativa no processo de construção, e tal inspeção e reaproveitamento podem se enquadrar neste passo de identificação de termos da ontologia. A fase destinada à integração de ontologias será elucidada na seção 6.2.8 desta proposta metodológica.

Tendo-se adquirido os termos do domínio através de princípios, métodos e técnicas[142] descritos anteriormente, o próximo passo seria registrar tais termos no documento Escopo da ontologia, introduzido na fase de especificação de requisitos da seção 6.2.3. Destaca-se, mais uma vez, que o escopo pode resultar, em princípio, num documento parcialmente completo em função de possíveis evoluções nos protótipos.

É válido destacar ainda que o fator cognitivo pode interferir no processo de aquisição de conhecimento, bem como em todo o processo de modelagem. Este se relaciona à subjetividade, ao conhecimento prévio, preferências, vieses e ideologias inerentes ao profissional *ontologista*. Tal

[141] A ferramenta DARE apóia uma abordagem semi-automática para construção de esquemas de classificação facetada. Informações sobre os algoritmos utilizados na ferramenta podem ser consultadas na referência.
[142] Dentre os métodos e técnicas apresentados, o *ontologista* precisa avaliar os mais adequados perante o contexto no qual se está modelando.

fator será determinante para a validação do conjunto consensual de termos obtido, isto é, assegurar que tal conjunto seja representativo do conhecimento sobre o domínio.

6.2.4.2 Definição de conceitos

A definição conceitual é importante na medida em que fornece as características do conceito que vão permitir seu agrupamento, indicar as relações, além de fornecer segurança ao *ontologista* no estabelecimento dos axiomas formais, caso o nível de formalidade da ontologia seja rigorosamente formal (os axiomas são discutidos na seção 6.2.4.12).

As características usadas na comparação entre os conceitos levam a uma classificação de conceitos (discutida na próxima seção) e é nesta perspectiva que se deve buscar a definição mais adequada, ou seja, aquela que atende aos propósitos da ontologia. Pode-se inferir daí que, se não se dominam os conceitos de uma determinada área de assunto, não se organiza uma ontologia de domínio. A explicação do conceito deve ficar clara ao *ontologista* e tal explicação ocorre no momento da definição do conceito.

Na atividade de definição são produzidas descrições textuais precisas e sem ambigüidades para os conceitos, as relações e os atributos representados pelos termos colhidos no passo apresentado na seção 6.2.4.1. Deve ser ressaltado que as descrições textuais não são feitas isoladamente, isto é, os termos não são definidos de forma independente como em um dicionário clássico, mas se definem uns em relação aos outros, conforme princípios da teoria da terminologia fundamentados na seção 2.1.2.

Os <u>princípios para a definição</u> de conceitos da ontologia nessa proposta metodológica são advindos da <u>teoria do conceito</u> e da <u>teoria da terminologia</u>, tendo em vista que Dahlberg posiciona a teoria geral da terminologia como precursora da teoria do conceito. Tais princípios seriam: a) identificação do objeto ou item de referência (descrita na seção 6.2.4.1); b) análise das características intrínsecas (do próprio objeto) e extrínsecas do objeto (funções do objeto ou relações com outros objetos), de forma a definir o conceito e as relações entre os conceitos, formando-se assim, sentenças corretas sobre o objeto; c) construção da taxonomia da ontologia a partir do princípio da contextualização, em que as definições dos conceitos e seus posicionamentos na estrutura estão diretamente ligados ao domínio ou área de conhecimento na qual a terminologia está sendo construída; e d) seleção dos termos como forma de expressar os conceitos presentes na ontologia.

Dahlberg (1978) em sua tríade discorre sobre as asserções corretas sobre um item de referência em que deve ser considerado o aspecto <u>intensional</u> e <u>extensional</u> para tal item, conforme foi fundamentado na seção 2.1.2 desta pesquisa. As ontologias possuem características intensional e

extensional, pois são constituídas de um vocabulário consensual formando uma rede conceitual de caráter intensional (GUARINO, 1998), além de possuírem extensões acerca de seus conceitos.

No processo de definição dos conceitos, o *ontologista* deve observar o nível intensional e o nível extensional do domínio, de forma a conceber a estrutura taxonômica, e deve também considerar: i) *definição genérica*, que determina o gênero próximo e a diferença específica do conceito; ii) *definição partitiva*, que é constituída pelas partes do conceito definido, ou seja, o todo e suas partes; e iii) *definição por extensão*, que lista os elementos ou o nome dos membros de um conceito através de exemplos de seu uso. Para tal, o processo de definição dos conceitos em nível intensional e extensional deve ser conduzido pelo *ontologista* e especialistas do domínio, pois somente quem conhece e entende do mesmo poderá fazer asserções corretas e verdadeiras sobre os conceitos envolvidos. Desse modo, a relação entre os níveis terminológico, intensional e extensional possibilitará a transformação de um conjunto de termos isolados em conceitos, relações e atributos que expressarão o vocabulário consensual do domínio.

No processo de definição das intensões é importante considerar a existência de termos ambíguos. Podem ocorrer as seguintes situações: um termo corresponde a apenas uma definição, o que não causa problemas; vários termos correspondem a uma definição, indicando a existência de termos sinônimos ou termos acrônimos[143], a partir dos quais é produzida uma *lista de sinônimos* bem como uma *lista de acrônimos*; e um termo corresponde a vários conceitos, o que configura ambigüidade. Algumas recomendações devem ser consideradas no tratamento de ambigüidades (USCHOLD e GRUNINGER, 1996): a) suspensão do uso do termo; b) esclarecimento da idéia relativa ao conceito, a partir de uma definição cuidadosa, e a utilização do menor número possível de termos técnicos; c) indicação dos conceitos importantes, justificando sua inserção na ontologia; e d) escolha de um único termo para o conceito. Uma técnica que pode ser utilizada na obtenção de um consenso acerca das noções intensionais e extensionais é a entrevista com especialistas, que podem ser fundamentadas em JAD.

Após todas as considerações elucidadas, o *ontologista* deve registrar o conhecimento adquirido no modelo glossário de termos que propõe os seguintes campos: *nome do termo*; *lista de sinônimos; lista de acrônimos* e *descrição*, que fornece o significado de um conceito.

[143] Palavra formada das primeiras letras ou sílabas de outras palavras, por exemplo, EMBRATUR (Empresa Brasileira de Turismo).

A Tabela 35 apresenta exemplos de termos em um domínio de viagem, que precisariam de seus conceitos definidos no campo *descrição*.

Nome do termo	Lista de sinônimos	Lista de acrônimos	Descrição
Ibéria Flight		IB Flight	
Viagem de trem	Viagem de estrada de ferro		
Viagem			
Pacote de viagem			
Número máximo de viajantes			
Data de chegada			
Data de partida			
Local de partida			
Disposto em			

Tabela 35 – Modelo de glossário de termos
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

6.2.4.3 Agrupamento de termos em categorias

Tendo-se definido os conceitos no glossário, o próximo passo é organizar tais conceitos em uma estrutura facetada a fim de organizar o conhecimento do domínio e, a partir de tal organização, buscar conhecer a natureza dos conceitos, na qual permitirá determinar com maior precisão as classes[144], as relações e os atributos da ontologia. Para tal determinam-se facetas cujas manifestações partem das categorias e subcategorias de Dahlberg (DAHLBERG, 1978) em sua teoria do conceito.

Para cada termo pertencente ao nível terminológico do domínio deve ser identificada a categoria na qual o termo pertence. Dahlberg (1978) determina a relação formal-categorial para classificar os conceitos da mesma natureza dentro de uma mesma categoria. Assim sendo, se o conceito possui componentes temporais, espaciais ou de posição (lugar), então a categoria será Dimensões; se o conceito compreende noções de operações, estados ou processos, então se pode dizer que a categoria será Atividades; se o conceito indicar propriedades, incluindo quantidades, qualidades ou relações, então a categoria será Propriedades; e se o conceito definir entidades, incluindo objetos materiais ou imateriais, então se pode dizer que a categoria fundamental será Entidades.

Um exercício prático advindo de Gomes et al. (1990, p.37) para se identificar as categorias relevantes em um domínio é o de indagar, por exemplo, qual o objeto de estudo e, a partir daí,

[144] Classe na teoria das ontologias é denominada conceito, conforme pôde ser percebido no decorrer do trabalho.

acompanhar a lógica do assunto. Assim, no domínio de **Viagem** que categorias poderiam ser identificadas? De modo genérico pode-se começar dizendo que a Viagem cuida do transporte de pessoas. Posta a lógica do assunto, foram identificadas duas categorias/subcategorias advindas de Dahlberg (DAHLBERG, 1978) para os termos transporte e pessoas:

(categoria/subcategoria)	**(termo)**
Atividade/processo	transporte
Entidade/objeto material	pessoas

Além das categorias, cada unidade terminológica deve ter um indicador de faceta, que seria uma indicação do princípio de divisão a ser utilizado na classificação facetada. Dentre os princípios de escolha das facetas e subfacetas são considerados o princípio da diferenciação e o princípio da exclusividade mútua, os quais são oriundos da teoria da classificação de Ranganathan.

O princípio da diferenciação é embasado em características de divisão em que as diferenças distinguem elementos da mesma classe. Por exemplo, a característica de divisão que pode ser usada na organização da faceta “Viagem de trem” seria “por fator econômico”. Sendo assim, os renques seriam “Trem econômico”, “Trem executivo” e “Trem de luxo”.

O princípio da exclusividade mútua determina que as facetas devem ser mutuamente exclusivas, assegurando a exclusividade das classes de um renque. Por exemplo, a faceta “Vôo da *British Airways*” possui três subfacetas, as quais utilizam o princípio da exclusividade mútua para assegurar que nenhum vôo possa ser simultaneamente um “vôo BA0068”, um “vôo BA0066” e um “vôo BA0069”.

Nesse estágio de trabalho intelectual, pode-se perceber o caráter subjetivo da atividade, uma vez que, a categorização e a classificação de termos implica na decisão muitas vezes subjetiva e passível de discussão, ou seja, as facetas podem ser construídas conforme perspectiva, pontos de vista ou dimensões de um domínio particular. Além disso, o estabelecimento de termos a serem usados no sistema e dos demais a serem empregados em caráter de remissiva envolvem um conhecimento amplo e atual do domínio ou área de assunto na qual se está modelando. Essa escolha também se relaciona com a *heteroglossia*, a linguagem própria dos integrantes da área.

Tendo-se elaborada a estrutura facetada, o *ontologista* terá uma maior facilidade em identificar os conceitos (classes), os atributos, as constantes e as relações do domínio, tendo-se em vista que os termos foram organizados em categorias envolvendo entidades e dimensões, que remetem a conceitos genéricos e comuns ao domínio; propriedades, que indicam atributos, constantes e relações; e atividades que indicam relações funcionais. Desse modo, o glossário de termos deve ser refinado no sentido de classificar os termos em: conceito, atributo, constante ou relação. A Tabela 36 apresenta o modelo glossário de termos com a adição do campo *Tipo* a fim de classificar os termos após a aplicação do método de análise facetada.

Nome do termo	Lista de sinônimos	Lista de acrônimos	Descrição	Tipo
Ibéria Flight				**Conceito**
Viagem de trem				**Conceito**
Viagem				**Conceito**
Pacote de Viagem				**Conceito**
Número máximo de viajantes				**Constante**
Data de chegada				**Atributo**
Data de partida				**Atributo**
Local de partida				**Relação**
Disposto em				**Relação**

Tabela 36 – Modelo glossário de termos refinado após a análise facetada

Recomenda-se verificar o conhecimento inserido dentro do glossário buscando: a) garantir a completude do conhecimento sobre os conceitos envolvidos; b) determinar a granularidade dos conceitos cobertos pela ontologia; e c) assegurar a ausência de redundância e imprecisão sobre nomes e descrições de conceitos.

6.2.4.4 Construção da estrutura taxonômica

Verificado se os conceitos estão reunidos por sua natureza, e após o glossário de termos ter sido refinado, o próximo passo consiste em verificar como os conceitos de mesma natureza se relacionam. Nesta forma de relacionamento, determinam-se as relações hierárquicas e as relações partitivas com o propósito de construir taxonomias dos termos classificados como conceitos no glossário. Segundo Dahlberg (1978), a relação hierárquica é de dois tipos: relação de abstração (relação gênero-espécie) e relação lateral (relação dos conceitos no renque). A relação partitiva existe entre o todo e suas partes, e as partes também podem estar relacionadas entre si.

A Figura 74 mostra um exemplo de taxonomia envolvendo conceitos de uma possível estrutura facetada elucidada na seção 6.2.4.3.

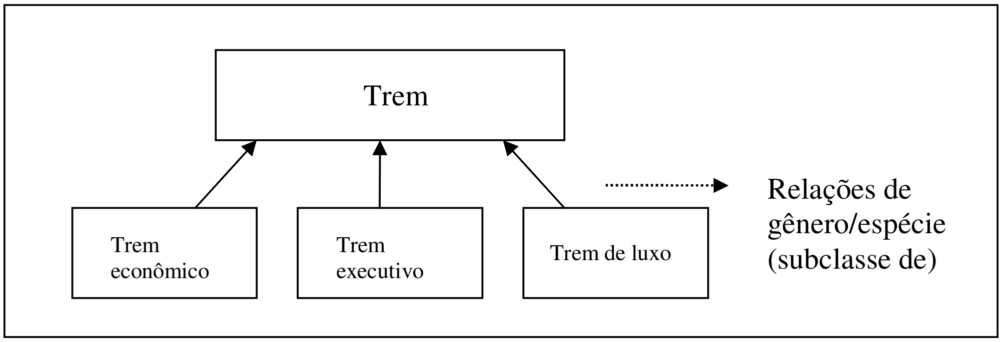


Figura 74 – Taxonomia da faceta Trem e suas subfacetas

A realização da atividade de organizar hierarquicamente os conceitos em uma estrutura taxonômica envolve uma combinação de métodos advindos da teoria das ontologias, denominados *top-down*, *bottom-up e middle-out*. Conforme já apresentado em algumas metodologias e métodos para construção de ontologias (capítulo 4), a abordagem *bottom-up* determina os conceitos de nível mais baixo na hierarquia para depois generalizá-los; a abordagem *top-down* determina os conceitos de nível mais alto na hierarquia para depois especializá-los; e a abordagem *middle-out* determina os conceitos centrais para então generalizá-los e especializá-los.

No processo *top-down*, o *ontologista* e especialistas do domínio trabalham juntos na identificação de conceitos-chave de alto-nível, a fim de conceber a camada superior da ontologia, na qual se representam conceitos genéricos e conceitos comuns ao domínio. Tal processo demanda criatividade por parte dos envolvidos, uma vez que ontologias de alto nível já existentes precisam ser analisadas e adequadas à conceitualização da ontologia construída. Desse modo, a abordagem *top-down* é vista como um processo informal e manual, ou seja, geralmente não se faz uso de recursos computacionais na determinação de categorias. Já os processos *bottom-up* e *middle-out* são conduzidos na elaboração da estrutura facetada do domínio em que questões de competência, suas respostas e outros tipos de documentos são analisados no propósito de identificar conceitos (específicos e centrais) a fim de organizá-los em facetas de alto nível, bem como em facetas do próprio domínio. Tais processos podem ser conduzidos de maneira sistemática, repetida e semi-automática como no exemplo de utilização da ferramenta DARE (PRIETO-DÍAZ, 2003), em que textos são analisados com o objetivo de extrair uma lista de termos-chave que denotem conceitos únicos; tais termos são agrupados em um editor de *clusters* que fornece um conjunto de *clusters* conceituais. O *ontologista* pode então modificar, agrupar e reagrupar tais *clusters* candidatos no intuito de determinar as facetas, que serão a base para o processo classificatório. Em tal processo, os termos são agrupados mediante propriedades similares resultando, assim, em uma estrutura facetada.

A estrutura facetada torna-se um instrumento útil no auxílio à comparação de categorias ou facetas sintetizadas na estrutura com os conceitos de ontologias de alto nível. Diante disso, a fase

de integração é viabilizada no sentido de que as facetas juntamente com seus focos possam ser comparadas mais facilmente com as categorias pertencentes a ontologias de alto nível. Tal comparação pode requerer modificações nas facetas bem como em seus focos na adequação da integração e, conseqüentemente, na ontologia que está sendo construída, constituindo-se assim de um processo iterativo. Logo, os conceitos advindos de ontologias de alto nível podem ser avaliados, modificados e validados através dos conceitos da estrutura facetada. Em suma, os termos são adaptados e dispostos adequadamente na nova estrutura taxonômica, abaixo de termos mais genéricos, obtidos em ontologias de alto nível, o que proporciona uma melhor capacidade de descrever o domínio. A Figura 75 ilustra os processos necessários para a construção da estrutura taxonômica da ontologia de domínio, em que são utilizados simultaneamente os métodos *top-down*, *bottom-up* e *middle-out*, além do método de análise facetada.

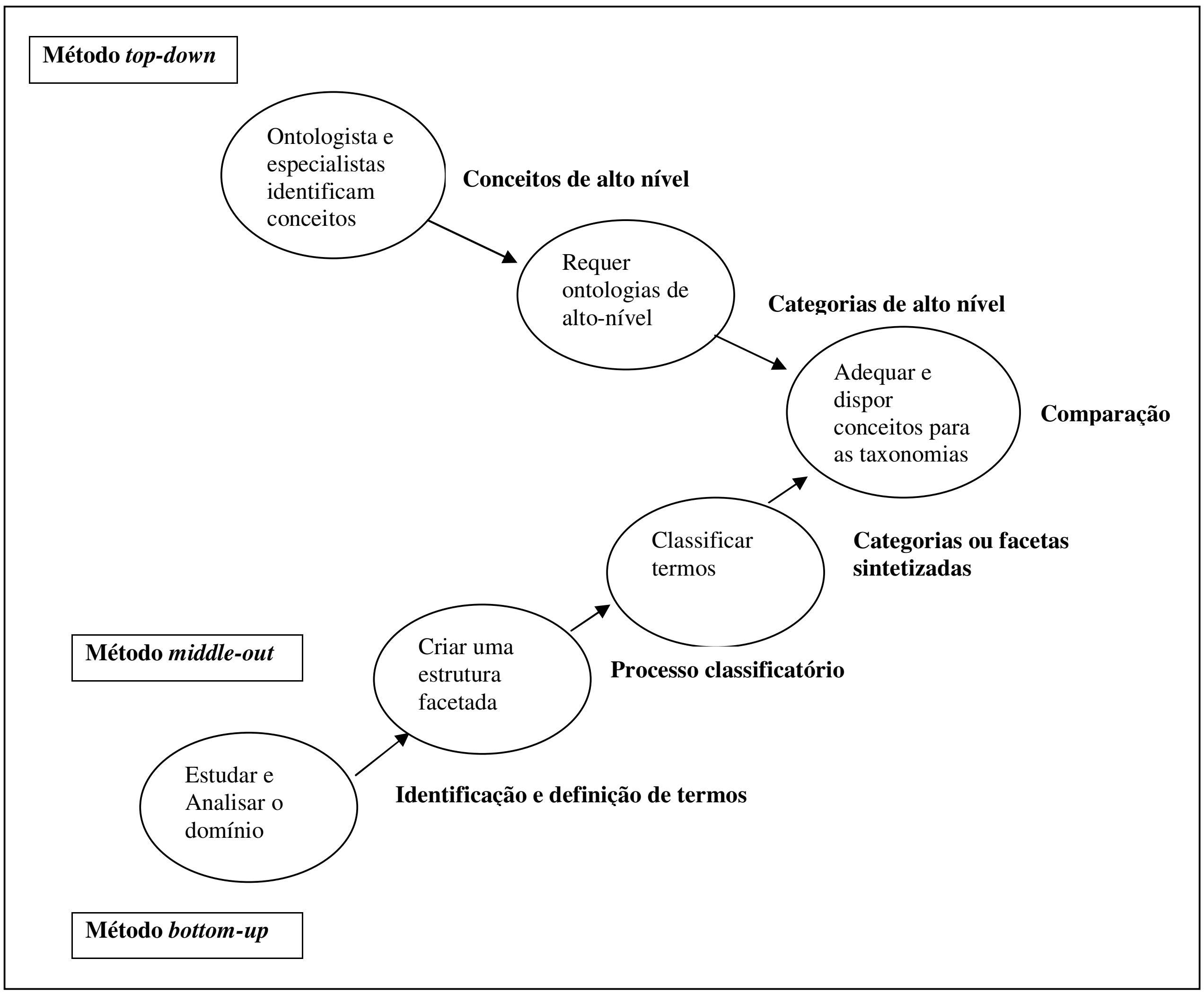


Figura 75 – Processos para construção da estrutura taxonômica

A Tabela 37 a seguir apresenta alguns princípios empíricos advindos do método 101 considerados relevantes por auxiliarem a construção de hierarquias de classes.

Princípios	Descrição
Conceber relação "é-um" ou "tipo-de"	A hierarquia de classe representa uma relação "é-um", portanto uma classe A é uma subclasse de B se toda instância de B é também uma instância de A. Ou, uma subclasse de uma classe representa um conceito que é um "tipo de" conceito que a superclasse representa.
Identificar transitividade na hierarquia de classes	Um relacionamento é considerado transitivo: se B é uma subclasse de A e C é uma subclasse de B, então C é uma subclasse de A.
Determinar nomes ou termos para as classes	Recomendação para definição dos termos para os conceitos: i) os nomes devem ser todos no singular (na prática, uma classe é mais usada no singular) ou todos no plural; ii) termos sinônimos para o mesmo conceito não representam classes diferentes. Apenas um termo deve estar associado a uma classe; iii) utilização de prefixo e sufixo nos termos a fim de diferenciar as classes das propriedades; iv) evitar abreviações nos termos dos conceitos; v) as subclasses deveriam incluir o nome de suas classes na definição de seus termos.
Identificar classes disjuntas	Classes são disjuntas se elas não podem ter instâncias em comum, além de não poderem ter nenhuma subclasse em comum.
Determinar quantidade de subclasses para uma classe	Não existe uma regra geral para o número de subclasses diretas que uma classe deve ter. Entretanto, muitas ontologias bem estruturadas têm entre duas e doze subclasses diretas. Algumas recomendações: i) se uma classe tem apenas uma subclasse direta pode existir um problema de modelagem ou a ontologia não está completa; ii) se uma classe tem mais do que doze subclasses, então categorias intermediárias podem ser necessárias.
Decidir pela inserção (ou não) de uma nova subclasse	Recomendações: subclasses de uma classe usualmente i) têm propriedades adicionais que a superclasse não tem; ou ii) têm restrições diferentes das superclasses; ou iii) participam de diferentes relacionamentos do que as superclasses.
Decidir por uma nova classe ou por um valor de propriedade	Recomendações a partir da questão: *de que maneira uma distinção de um conceito é considerada importante no domínio?* Se as distinções sobre o conceito não implicarem diretamente em suas relações com outros objetos, então não há necessidade de introduzir uma classe para tal conceito e sim criar uma propriedade para a classe envolvida com o mesmo. Em contrapartida, se uma distinção é importante no domínio e implicar em distinções para diferentes objetos e relações, então há necessidade de criar uma nova classe para a distinção.
Decidir por uma instância ou por uma classe	Instâncias individuais são conceitos mais específicos representados na base de conhecimento. Para tal decisão, recomenda-se analisar as questões de competência levantadas na fase de especificação de requisitos e identificar os conceitos mais específicos que constituirão respostas a tais questões. Tais conceitos podem ser bons candidatos para instâncias na base de conhecimento. Por outro lado, se conceitos possuem propriedades comuns e formam uma hierarquia natural, então se deve representá-los como classes que acomodarão instâncias individuais com tais propriedades.

Tabela 37 – Princípios empíricos para construção de hierarquias de classes

Recomenda-se ainda verificar o conhecimento inserido dentro das taxonomias de forma a visualizar graficamente os conceitos do glossário de termos tornando-se possível assegurar: a) que nenhum conceito seja repetido; e b) nenhuma taxonomia isolada em relação aos conceitos relacionados. Como verificação cruzada entre representações intermediárias recomenda-se observar se todos os conceitos mostrados nas taxonomias estão definidos no glossário de termos. Caso ocorra algum desencontro de informação, o conceito deve ser adicionado no glossário ou removido da taxonomia.

6.2.4.5 Construção do Diagrama de Relações Binárias

Uma vez que a estrutura taxonômica tenha sido criada e avaliada, o próximo passo é construir diagramas de relações binárias que objetivam estabelecer relacionamentos entre conceitos da mesma taxonomia ou mesmo de taxonomias diferentes.

Estabelecer relações entre os conceitos significa posicioná-los no sistema de conceitos, ou seja, estabelecer-lhes gênero, espécie e as associações com os outros conceitos. Isso é possível quando efetivamente se domina seu significado, isto é, quando são conhecidas as características relevantes para o sistema. O esquema facetado contribui em tal perspectiva quando evidencia a natureza dos conceitos e de suas relações.

A Figura 76 apresenta uma porção de um diagrama de relações entre conceitos no exemplo do domínio de viagem. As relações “local de partida” e “local de chegada ” e as relações inversas “é local de chegada de” e “é local de partida de” se conectam aos conceitos de alto nível “Viagem” e “Local” de suas respectivas taxonomias de viagens e locais. Desse modo, as classes referentes aos domínios e aos escopos, bem como suas relações ou propriedades, podem ser identificados precisamente e delimitados no momento da implementação em linguagens como RDFS e OWL (ver seções 2.3.3.4.3 e 2.3.3.4.4).

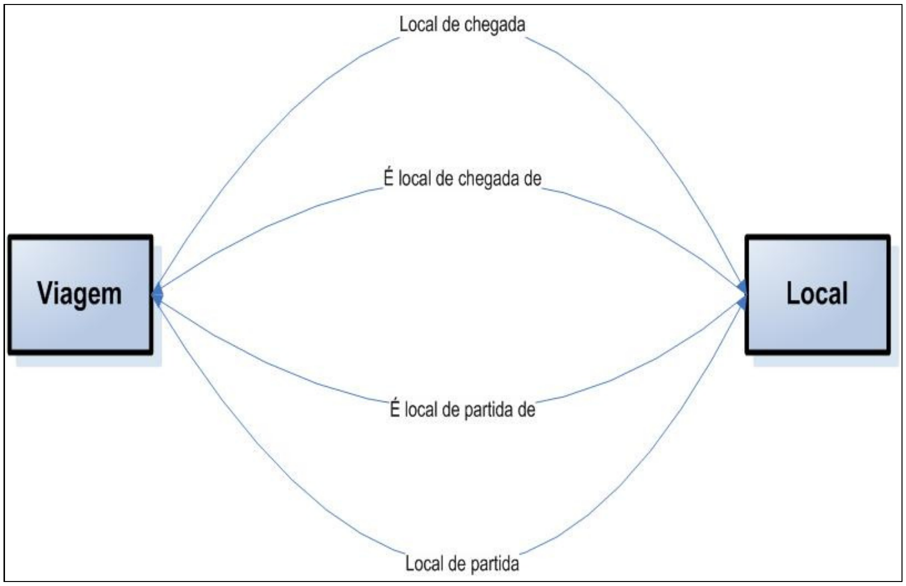


Figura 76 – Exemplo de diagrama de relações
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004, p.136)

6.2.4.6 Construção do dicionário de conceitos

Tendo-se desenvolvido as taxonomias de conceitos e os diagramas de relações, a proposta metodológica propõe construir um dicionário de conceitos para especificar as propriedades que descrevem cada conceito da taxonomia, incluindo suas relações, suas instâncias, seus atributos de classe e atributos de instância. A Tabela 38 apresenta a proposta de um modelo para dicionário de conceitos.

Nome do Conceito	Atributos de classe	Atributos de instância	Relações
AA7462	Nome da companhia	Data de partida Data de chegada	
American Airlines Flight	Nome da companhia	Data de partida Data de chegada	
Hotel Cinco Estrelas	Número de estrelas		
Local		Nome Extensão	É local de partida de É local de chegada de
Alojamento		Preço de quarto padrão	Disposto em
Viagem		Data de chegada Nome da companhia Data de partida	Local de partida Local de chegada
Pacote de viagem		Orçamento Preço final Nome Número de dias Restrições da viagem	Local de chegada Local de partida Viagens para Acomodado em

Tabela 38 – Modelo dicionário de conceitos
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Destaca-se que as relações especificadas para cada conceito são aquelas cujo domínio é o próprio conceito. Por exemplo, o conceito “Viagem” possui duas relações “local de partida” e “local

de chegada". Além disso, podem ocorrer repetições de relações, de atributos de classes e de atributos de instância em diferentes conceitos pela razão de tais elementos serem declarações para conceitos. Tal fato pode ser constatado nos exemplos da Tabela 38.

Recomenda-se verificar o conhecimento inserido dentro do dicionário buscando: a) garantir a completude do conhecimento sobre os conceitos envolvidos; e b) a consistência sobre as relações e os atributos de classes e instâncias, em que devem fazer sentido para os conceitos envolvidos.

Uma vez que o dicionário de conceitos tenha sido construído e avaliado, recomenda-se que o *ontologista* em conjunto com especialistas do domínio descreva em detalhe as relações binárias, os atributos de instâncias e os atributos de classes identificados no dicionário de conceitos. Além disso, as constantes identificadas no glossário de termos também devem ser detalhadas em representação específica. Desse modo, o processo metódico de construção do conhecimento acerca da conceitualização da ontologia pode ser conduzido de forma eficiente em direção a uma implementação bem sucedida.

6.2.4.7 Descrição de relações binárias

Conforme exposto, recomenda-se criar uma representação intermediária para cada relação explicitada no dicionário de conceitos. Tal representação contém os seguintes campos: *conceito ou classe de origem*, *nome da relação*, *conceito ou classe destino*, *cardinalidade* e *relação inversa*. A Tabela 39 ilustra o modelo proposto para a representação intermediária em questão.

Conceito origem	Nome da Relação	Conceito destino	Cardinalidade	Relação inversa
Alojamento	**Disposto em**	Local	1	
Pacote de viagem	**Acomodado em**	Alojamento	N	
Viagem	**Local de chegada**	Local	1	É local de chegada de
Viagem	**Local de partida**	Local	1	É local de partida de
Pacote de viagem	**Local de chegada**	Local	1	É local de chegada de
Pacote de viagem	**Local de partida**	Local	1	É local de partida de

Tabela 39 – Modelo para detalhamento de relações binárias
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

6.2.4.8 Descrição de atributos de instância

Esta tarefa demanda detalhar os atributos de instância especificados no dicionário de conceitos. Os atributos de instância são atributos cujos valores podem ser diferentes para cada

instância de um conceito. O detalhamento é direcionado a um modelo contendo os seguintes campos: *nome do atributo de instância*; *nome do conceito*, no qual o atributo pertence; *tipo de valor* (ex: *float*, *boolean*, *string*), que determina a restrição no momento do preenchimento do campo na classe; *cardinalidade máxima e mínima*, que especifica o número de valores que o atributo pode receber; *atributos de instância*, *atributos de classe* e *constantes usados para inferir* valores do atributo em questão; *regras* ou fórmulas, que faz referência cruzada com a tabela de regras (será discutida na seção 6.2.5.2) e permite inferir o valor para o atributo; *atributos inferidos*, que contêm o nome dos atributos de instância cujos valores podem ser inferidos a partir do atributo em questão; e *referência* para registrar a fonte de informação (relatório, especialista, reunião, ontologias) usada no detalhamento do atributo. Alguns atributos foram omitidos das Tabelas 40, 41 e 42 por questão de espaço. São eles: unidade de medida para valores numéricos; precisão para valor numérico; faixa de valores para possíveis valores no atributo; e valores padrão, caso sejam conhecidos.

Nome do atributo de instância	**Nome do conceito**	**Tipo de valor**	**Card. (máx. e min.)**	**Atrib. de instância usados para inferir**	**Atrib. de classe usados para inferir**	**Const. usadas para inferir**	**Regra**	**Atrib. inferidos**	**Ref.**
Orçamento	Viagem executiva	Float	(0,1)						Reunião Alfa
Nome	Local	String	(1,N)						
Extensão	Local	Integer	(1,1)						
Preço de quarto padrão	Alojamento	Float	(0,1)						Relatório Beta
Data de chegada	Viagem	Date	(0,1)						Ontologia de Viagem
Nome da companhia	Viagem	String	(0,N)						
Data de partida	Viagem	Date	(0,1)						Ontologia de Viagem
Orçamento	Pacote de viagem	Float	(0,1)					Preço Final	
Preço final	Pacote de viagem	Float	(0,1)	Orçamento					

Tabela 40 – Modelo para detalhamento de atributos de instância
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar o conhecimento dentro da tabela de atributos de instância de forma a garantir a consistência do conhecimento de senso comum para a ontologia construída. A saber: a) verificar a correspondência entre valores preenchidos nas linhas da tabela de atributos de instância. Por exemplo: em uma **linha x** da tabela, quando o campo *atributos de instância usados para inferir* for preenchido, a linha correspondente ao atributo indicado neste último campo deve possuir em seu

campo *atributos inferidos* o nome do atributo representado na **linha x**; b) valores padrão devem ser preenchidos de acordo com o tipo de dado definido para o atributo. No que diz respeito à verificação do conhecimento entre representações intermediárias tem-se: a) verificar a correspondência entre os atributos de instância detalhados e os atributos de instância listados no dicionário de conceitos; b) verificar a correspondência de valores preenchidos nos campos *atributos de classe usados para inferir* e *constantes usadas para inferir* com o valor preenchido no campo *atributos de instância inferidos* das tabelas de atributos de classe e constantes.

6.2.4.9 Descrição de atributos de classe

Para esta tarefa, um modelo é proposto para detalhar todos os atributos de classe inseridos no dicionário de conceitos. O modelo especifica os seguintes campos: *nome do atributo de classe*; *conceito em que o atributo é definido*, que indica a classe conceitual para o atributo; *tipo de valor*; *valor do atributo* de classe; *cardinalidade máxima e mínima*; *atributos de instância inferidos*, que contém o nome dos atributos de instância cujos valores podem ser inferidos a partir do valor do atributo de classe em questão; e *referência* para registrar a fonte de conhecimento. A Tabela 41 ilustra exemplos de atributos de classes com seus detalhes.

Nome do atributo de classe	Conceito em que o atributo é definido	Tipo de valor	Valor do atributo	Cardinalidade (máx. e min.)	Atributos de instância inferidos	Referência
Nome da companhia	American Airlines Flight	String	AAF	(1,1)		
Nome da companhia	AA7462	String	AA	(1,1)		
Número de estrelas	Hotel Cinco Estrelas	Integer	5	(1,1)		
Número de estrelas	Hotel Quatro Estrelas	Integer	4	(1,1)		

Tabela 41 – Modelo para detalhamento de atributos de classe
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar o conhecimento entre representações intermediárias como: para todos os atributos de classe identificados no dicionário de conceitos deve existir uma linha na tabela de atributos de classe, e para cada linha na tabela de atributos de classe deve existir um conceito no dicionário.

6.2.4.10 Descrição de constantes

Para cada constante definida no glossário de termos deve-se descrevê-la em detalhe a partir dos seguintes campos: *nome da constante*; *tipo de valor*; *valor* da constante; *atributos de instância inferidos*, que relaciona os atributos de instância cujos valores podem ser inferidos a partir do valor da constante em questão; e *referência* das fontes de conhecimento, nas quais as constantes foram identificadas. A Tabela 42 apresenta a proposta de um modelo para detalhamento de constantes.

Nome da Constante	Tipode valor	Valor	Atributos de instância inferidos	Referência
Número máximo de viajantes	Integer	200		

Tabela 42 – Modelo para detalhamento de constantes
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar o conhecimento dentro da tabela de constantes: se uma constante aparecer na tabela de atributos de instância ou de regras, deve também ser inserida na tabela de constantes.

6.2.4.11 Definição de instâncias

Instâncias são valores específicos de uma classe ou conceito e são conhecidas a partir das extensões levantadas sobre os conceitos de um domínio. Esta tarefa consiste em definir as instâncias relevantes para os conceitos e direcioná-las a uma representação intermediária específica. A Tabela 43 apresenta a proposta de um modelo com exemplos de instâncias do conceito "AA7462". Os campos a serem preenchidos no modelo são: *nome da instância*; *nome do conceito* ou classe que a instância pertence; *atributos* para preenchimento na instância em função das classes em que as instâncias estão ligadas; *valores* desses atributos, que podem ser numérico, conjunto de valores ou "desconhecido".

Nome da Instância	Nome do Conceito	Atributos	Valores
AA7462_Feb08_2002	AA7462	Nome da companhia Data de partida Data de chegada	American Airlines 05/09/2007 05/09/2007
AA7462_Feb16_2002	AA7462	Nome da companhia Data de partida Data de chegada	American Airlines 30/11/2007 30/11/2007

Tabela 43 – Modelo para definição de instâncias
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar o conhecimento dentro da tabela de instância: a) consistência entre atributos e seus valores; b) ler a descrição da instância de maneira cuidadosa para evitar redundâncias entre instâncias. Quanto à verificação do conhecimento entre representações

intermediárias: a) para todas as instâncias identificadas no dicionário de conceitos existe uma linha na tabela de instância correspondente e vice-versa; b) manter consistência entre os valores de atributo da instância e o campo *tipo de valor* definido em sua tabela de atributos de instância; c) manter consistência entre o número de valores declarado no atributo e o campo *cardinalidade* definido em sua tabela de atributos de instância; d) observar a propagação de atributos usando herança de classes, ou seja, a instância seria preenchida com valores concretos de acordo com a hierarquia definida na taxonomia.

6.2.4.12 Considerações sobre a fase de modelagem conceitual

Percebe-se neste ponto a semelhança entre algumas atividades de conceitualização, da fase de modelagem conceitual, e o processo de modelagem orientada a objetos. Tal semelhança estaria principalmente na concepção da estrutura taxonômica da ontologia, bem como nos diagramas de relações. Ambas as representações intermediárias necessitam fazer associações semânticas relacionadas à generalização, à agregação e à associação entre classes. Conforme pôde ser visto na fundamentação teórica (seção 2.3.2.2.1), a UML possui construtos abstratos o suficiente para permitir a representação de ontologias, como classes, atributos e relacionamentos de generalização, associação e agregação (OMG, 2003). E, ainda, permite restrições nos objetos através da linguagem auxiliar OCL, utilizada para expressar regras em UML.

O potencial dos padrões orientados a objeto pode ser usado na modelagem de ontologias através do diagrama de classe. A UML tem uma ampla e rápida expansão de uma comunidade de usuários, sendo que estes usuários provavelmente estarão mais familiarizados com a notação UML do que lógica descritiva, por exemplo. Ao contrário do formalismo da lógica descritiva, há um padrão de representação gráfica para modelos expressados em UML. Dessa maneira, uma representação gráfica seria importante, por exemplo, em permitir usuários de sistemas de informação pesquisarem uma ontologia e descobrir conceitos que podem aparecer nas consultas (CRANEFIELD e PURVIS, 1999). A representação gráfica torna-se, portanto, importante na medida em que viabiliza a comunicação entre as pessoas envolvidas nos sistemas de informações.

Diante a tal constatação, sugere-se a utilização do diagrama de classes da UML nas atividades de modelagem conceitual da ontologia, bem como da OCL para explicitação de restrições sobre os objetos das classes. Desse modo, o *ontologista* responsável pela formalização e implementação do modelo conceitual terá maior facilidade no entendimento da modelagem do domínio, resultando, assim, em uma especificação mais próxima da realidade.

Assim como tem crescido progressivamente o uso de UML, tem aumentado também um vasto conjunto de ferramentas CASE (do inglês *Computer-Aided Software Engineering*), as quais

suportam esta linguagem por possuir mecanismos que permitem sua própria extensão para diversos domínios de problema. As mais conhecidas são o *ArgoUML*[145], o *Poseidon* e o *Unisys Rose XML Tools*[146].

Finalmente, a Tabela 44 apresenta a relação dos produtos gerados na fase de modelagem conceitual, os quais serão avaliados e destinados às fases seguintes: formalização do modelo conceitual, concluindo, portanto, a atividade de conceitualização, e implementação do modelo.

Informação de saída	Destino	
	Fase	Atividade
Estrutura taxonômica	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Diagrama de relações	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Dicionário de conceitos	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Descrição das relações binárias	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Descrição dos atributos de instância	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Descrição dos atributos de classe	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Descrição das constantes	Formalização	Formalização das questões
	Implementação	Implementação da ontologia
Definição das instâncias	Implementação	Implementação da ontologia
Escopo da ontologia	Formalização	Formalização das questões
Todas as representações intermediárias	Avaliação	Avaliação de resultados
Categorias ou facetas sintetizadas	Integração	Reuso de ontologias

Tabela 44 – Produtos resultantes da fase de modelagem conceitual

6.2.5 Fase de Formalização

Para ontologias que demandam um nível de rigor formal, o *ontologista* deve levar em consideração a fase de formalização. Uma vez que a estrutura da ontologia (conceitos, estrutura taxonômica, relações, atributos e constantes) tenha sido definida através da atividade de conceitualização, cabe ao *ontologista* descrever precisamente as questões de competência, os axiomas e as regras em uma linguagem de representação do conhecimento como, por exemplo, a lógica de primeira ordem e a lógica descritiva (ver seção 2.3.1.3). Tal tarefa é recomendada para assegurar que o formalismo da ontologia seja verificado (fase de avaliação) antes mesmo de sua implementação, buscando, assim, a consistência da ontologia frente às questões de competência. Além disso, a fase de formalização permite ao *ontologista* avaliar se a linguagem de implementação escolhida é compatível com o formalismo lógico usado na formalização da terminologia. Desse

[145] http://argouml.tigris.org/
[146] http://www.cin.ufpe.br/~in1006/2003/PoseidonRose.ppt

modo, torna-se possível verificar se a linguagem de implementação é capaz de fazer o tratamento relacionado à representação do domínio da mesma forma como a linguagem de representação formal o faria.

A Tabela 45 apresenta os produtos necessários à realização da fase de formalização.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Questões de competência	Especificação de requisitos	Elaboração das questões de competência
Escopo da ontologia	Modelagem conceitual	Identificação de termos
Estrutura taxonômica	Modelagem conceitual	Conceitualização
Diagrama de relações	Modelagem conceitual	Conceitualização
Dicionário de conceitos	Modelagem conceitual	Conceitualização
Descrição das relações binárias	Modelagem conceitual	Conceitualização
Descrição dos atributos de instância	Modelagem conceitual	Conceitualização
Descrição dos atributos de classe	Modelagem conceitual	Conceitualização
Descrição das constantes	Modelagem conceitual	Conceitualização
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 45 – Produtos requeridos na fase de formalização

As seções a seguir se encarregam de apresentar a formalização das questões de competência, dos axiomas e das regras.

6.2.5.1 Formalização das questões de competência

As questões de competência são definidas formalmente como um vínculo de problemas consistentes em relação aos axiomas na ontologia (GRUNINGER e FOX, 1995), ou seja, qualquer solução para a questão de competência deve estar vinculada com os axiomas formais. Para tal, criam-se regras descritas em linguagem formal, a fim de definir semanticamente as sentenças envolvidas nas questões.

No exemplo fornecido na seção 6.2.3 da fase de especificação de requisitos, a questão de competência informal poderia ser representada em lógica de primeira ordem de acordo com a terminologia definida na fase de modelagem conceitual. Um exemplo do modelo proposto para descrição de questões de competência formal é apresentado na Tabela 46. As constantes significam: *c* é o viajante, *b* é a quantidade máxima que o viajante pode gastar, *a* é a idade do viajante, *e* são suas preferências; e as variáveis x e y significam respectivamente: destino da viagem e informação

sobre a viagem. Percebe-se no exemplo que a questão formal faz menção apenas a conceitos e a relações, permanecendo em branco os atributos referenciados.

Nome da questão	Destinos mais apropriados para uma viagem
Descrição em linguagem natural	Considerando as preferências de um viajante (viagem por cultura, para as montanhas, para a praia, etc) e algumas restrições (econômica ou sobre a própria viagem), quais destinos são mais apropriados?
Expressão formal	$\exists \$x\ \$y$ (destino($x) $\wedge$ informação_de_viagem($y) $\wedge$ quer_viajar(c,$y) $\wedge$ idade(c, a) $\wedge$ preferencias (c, e) $\wedge$ despesa_maxima(c, b) $\wedge$ tem_destino($y,$x))?
Conceitos referenciados	destino, informação_de_viagem
Atributos referenciados	----
Relações referenciadas	quer_viajar, idade, preferencias, despesa_maxima, tem_destino
Variáveis	$x $y

Tabela 46 – Modelo para descrição de questões de competência formal
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar se todos os termos nas sentenças das questões de competência formal estão incluídos na terminologia da ontologia, de forma a estabelecer um vínculo entre a ontologia e as questões que ela precisará resolver.

6.2.5.2 Formalização dos axiomas e das regras

Os axiomas formais e as regras ou fórmulas são especificados em uma linguagem de representação do conhecimento. Conforme já visto na fundamentação teórica e na análise das metodologias e métodos para construção de ontologias, os axiomas são expressões lógicas que representam verdades sobre os objetos do domínio, e buscam especificar restrições na ontologia. Regras são geralmente usadas para inferir conhecimento na ontologia a partir de valores de atributos, instâncias, relações, etc.

Para facilitar o processo de definição dos axiomas, considerado o mais difícil dentre os processos de construção de ontologias (GRUNINGER e FOX, 1995), a definição é guiada pelas questões de competência formal previamente definidas. Assim, os axiomas na ontologia devem ser necessários e suficientes para expressar as questões de competência e caracterizar as suas soluções. Finalmente, para apoiar o entendimento do axioma, recomenda-se fornecer uma descrição em linguagem natural. A Tabela 47 apresenta a proposta de um modelo para a descrição de axiomas formais, no qual ilustra exemplos em lógica de primeira ordem de um axioma que restringe a logística de uma viagem de trem na Europa. As variáveis usadas na expressão lógica são: ?X para "Viagem de trem"; ?Y para "Local de partida"; e ?Z para "Local de chegada".

Nome do axioma	Trem dentro da Europa
Descrição em linguagem natural	Todo trem que parte de uma localização na Europa deve chegar à outra localização na Europa.
Expressão formal	∀ (?X,?Y,?Z) ([Viagem de trem](?X) and [Local de partida] (?X,?Y) and [Local de chegada] (?X,?Z) and [Local Europa] (?Y) →[Local Europa](?Z))
Conceitos referenciados	Viagem de trem Local Europa
Atributos referenciados	-----
Relações referenciadas	Local de partida Local de chegada
Variáveis	?X ?Y ?Z

Tabela 47 – Modelo para descrição de axiomas formais
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar se os conceitos referenciados possuem as propriedades (atributos e relações) explicitadas na expressão formal. Por exemplo, no dicionário de conceitos, a relação "Local de partida" foi vinculada ao conceito "Viagem" (ver tabela 38). Desse modo, a restrição envolvendo as instâncias do conceito "Viagem de trem" será precisa somente se este conceito for uma subclasse do conceito "Viagem".

A Tabela 48 apresenta um exemplo de uma regra usando o modelo <*condição*> *então* <*conseqüência*>. A regra infere o nome da companhia de um navio, dado que o mesmo parte da Europa. As variáveis usadas na expressão lógica são: ?X para "Navio" e ?Y para "Local Europa".

Nome da regra	Regra da companhia Alfa
Descrição em linguagem natural	Todo navio que parte da Europa é preparado pela companhia Alfa.
Expressão formal	If [Local Europa] (?Y) and Navio(?X) and [Local de partida] (?X, ?Y) then [Nome da companhia] (?X, "Alfa")
Conceitos referenciados	Navio Local Europa
Atributos referenciados	Nome da companhia
Constantes referenciadas	--------
Relações referenciadas	Local de partida
Variáveis	?X ?Y

Tabela 48 – Modelo para definição de regras
Fonte: Adaptado de Fernandez, Gómez-Perez e Corcho (2004)

Recomenda-se verificar o conhecimento entre representações intermediárias: a) verificar se o campo *regra* especificado na tabela de atributos de instância corresponde à tabela de regra referenciada; b) verificar na tabela de regras, se os campos *atributos referenciados*, *conceitos referenciados* e *relações referenciadas* encontram-se preenchidos de acordo com suas tabelas correspondentes; c) verificar se o conteúdo dos campos *atributos de instância*, *atributos de classe* e

constantes usados para inferir valores de um atributo corresponde com o conteúdo do campo *expressão formal* da tabela de regras.

A Tabela 49 mostra os produtos resultantes da fase de formalização e as fases e atividades que necessitam de tais produtos. Os axiomas e regras formais especificados a partir das questões de competência formal são direcionados à fase de implementação da ontologia. O campo "formalizado por", localizado no documento **Escopo da ontologia,** é preenchido com o nome da pessoa responsável pela formalização e direcionado também à fase seguinte.

Informação de saída	Destino	
	Fase	Atividade
Axiomas e regras formais	Avaliação	Avaliação do formalismo
	Implementação	Implementação da ontologia
Escopo da ontologia	Implementação	Implementação da ontologia

Tabela 49 – Produtos resultantes da fase de formalização

6.2.6 Fase de Implementação

Para a realização da última fase do processo de desenvolvimento, os produtos listados na Tabela 50 devem ser considerados como entrada, de forma que a fase de implementação possa ser conduzida de maneira eficiente, possibilitando, assim, que a ontologia possa ser bem construída.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Escopo da ontologia	Formalização	Formalização
Estrutura taxonômica	Modelagem conceitual	Conceitualização
Diagrama de relações	Modelagem conceitual	Conceitualização
Dicionário de conceitos	Modelagem conceitual	Conceitualização
Descrição das relações binárias	Modelagem conceitual	Conceitualização
Descrição dos atributos de instância	Modelagem conceitual	Conceitualização
Descrição dos atributos de classe	Modelagem conceitual	Conceitualização
Descrição das constantes	Modelagem conceitual	Conceitualização
Definição das instâncias	Modelagem conceitual	Conceitualização
Axiomas e regras formais	Formalização	Formalização dos axiomas e das regras
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 50 - Produtos requeridos na fase de implementação

As seções a seguir se encarregam de apresentar as atividades necessárias ao processo de implementação de ontologias.

6.2.6.1 Escolha da ferramenta e da linguagem de implementação

A fase de implementação objetiva produzir modelos computáveis através de uma linguagem legível para a máquina, que seja capaz de atender aos compromissos ontológicos estabelecidos com a conceitualização. Desse modo, esta linguagem deve considerar axiomas em suas primitivas, de modo a fornecer semântica formal para o vocabulário ou terminologia do domínio, restringindo, assim, as suas possibilidades interpretativas.

O objetivo principal da linguagem é, portanto, permitir que um *ontologista*, usando uma ferramenta, possa representar quaisquer axiomas pertinentes à estrutura modelada e fazer consultas à ontologia, de modo a avaliar se a mesma satisfaz às questões de competência colocadas como requisito. Para tal, recomenda-se a utilização de uma linguagem para a descrição de ontologias que possua um mecanismo de inferência associado.

Na fundamentação teórica da presente pesquisa foram apresentadas algumas linguagens de representação de ontologias (ver seção 2.3.3.4) baseadas em XML, como RDF, RDFS e OWL. Tais linguagens são consideradas padrões bem aceitos na Web, porém possuem diferenças no que diz respeito ao poder de expressividade na definição semântica da terminologia.

Nesta proposta metodológica para construção de ontologias entende-se que a *Ontology Web Language* – OWL é a mais indicada pelas seguintes razões: i) considerada padrão W3C para descrição de ontologias, viabilizando e estendendo as possibilidades de integração com ontologias implementadas em padrão Web; ii) trabalha com axiomas; iii) disponibiliza mecanismos de inferência, que permitem submeter a ontologia à avaliação; e iv) possui suporte da ferramenta *Protégé*, que facilita a utilização da linguagem e conseqüentemente a implementação da ontologia.

A ferramenta *Protégé* é recomendada nessa proposta metodológica pelas seguintes razões: i) possui interface amigável; ii) possibilita documentar objetos, conforme pôde ser constatado na análise do método 101; iii) é de fácil obtenção por ser de domínio público; iv) possui uma comunidade de pesquisados que contribui com a sua evolução, mantendo a ferramenta sempre atualizada; e v) possui documentação disponível.

Em pesquisa recente envolvendo o universo acadêmico e da indústria, constatou-se que mais de 75% dos *ontologistas* têm selecionado a OWL para o desenvolvimento de ontologias; e 68% optaram pelo uso do *Protégé* (CARDOSO, 2007, pgs. 4-5).

6.2.6.2 Construção de classes, propriedades e restrições

Uma vez determinada a ferramenta e a linguagem para implementação, o próximo passo é organizar as representações intermediárias concebidas na fase de modelagem, bem como os

axiomas e regras concebidos na fase de formalização, a fim de serem utilizadas no processo de construção de classes, propriedades e restrições.

Em um projeto de construção de ontologias, a equipe envolvida pode ser interdisciplinar, sendo possível que em cada fase trabalhem profissionais com competências distintas e complementares. Diante disso, o profissional com habilidade na linguagem usada na formalização do modelo conceitual pode mapear as questões de competência (no formato de axiomas) e regras para os construtores de classe e axiomas da OWL ou da linguagem escolhida. Caso contrário, o profissional pode mapear a terminologia, concebida através das representações intermediárias (ver Tabela 50), para os construtores e axiomas da linguagem de implementação. Deste modo, os axiomas da linguagem são associados aos conceitos, atributos e relações da ontologia, os quais foram definidos nas representações intermediárias e através da formalização das questões de competência, cuja elaboração é feita a partir de tais elementos. É válido ressaltar que os tipos de relações discutidas em toda a pesquisa (na elucidação da teoria da terminologia, da teoria do conceito e da teoria das ontologias) como relações de subclasse, interseção, disjunção, equivalência, funcional, inversa, transitiva, dentre outras, são suportadas pelos construtores de classes e axiomas da OWL, conforme exibição na seção 2.3.3.4.4 (figuras 40 e 41).

E, finalmente, o *Protégé* utiliza os recursos de *slots* e facetas, respectivamente para definir propriedades (atributos e relações) das classes e restrições sobre tais propriedades, conforme pôde ser visto na análise do método 101 (ver seção 4.7).

6.2.6.3 Criação de instâncias de conceitos

Uma vez que as classes, seus atributos, suas relações e suas restrições tenham sido criadas através dos construtores e axiomas da linguagem, o próximo passo é criar as instâncias de tais classes ou conceitos, que também foram previamente definidas pela representação intermediária “definição das instâncias” na fase de modelagem conceitual. Caso seja necessário especificar alguma restrição sobre a instanciação de conceitos, axiomas precisam ser definidos sobre tais objetos.

Ao final da tarefa de implementação, a ontologia é avaliada a fim de ser validada perante as questões de sua competência. Além disso, o campo “implementado por”, localizado no documento **Escopo da ontologia,** é preenchido com o nome da pessoa responsável pela implementação e é finalmente dado como concluído. Tal documento é destinado à fase de gerenciamento do projeto, em que será submetido ao controle de versão dos protótipos da ontologia.

A Tabela 51 apresenta a relação dos produtos resultantes da fase e seus destinos.

Informação de saída	Destino	
	Fase	Atividade
Ontologia implementada	Avaliação	Validação da ontologia
Escopo da ontologia	Gerenciamento do projeto	Gerenciamento do protótipo

Tabela 51 - Produtos resultantes da fase de implementação

6.2.7 Fase de Manutenção

Uma vez que o processo de desenvolvimento tenha sido finalizado e avaliado, alguns ajustes e correções em função de eventuais problemas mapeados na fase de avaliação da ontologia podem ser necessários. Nesse contexto, o processo pós-desenvolvimento, denominado manutenção, executa os tratamentos necessários e corretos perante a necessidade identificada. A Tabela 52 apresenta a relação de produtos necessários à fase no sentido da mesma ser realizada. O fluxo de informação é o seguinte: uma determinada fase produz informação ou produto e o envia para avaliação. Esta executa a verificação e, caso seja necessário, reporta o resultado à manutenção. Esta, por sua vez, executa as tarefas cabíveis e retorna o resultado à fase de gerenciamento do projeto, recomendando melhorias nos próximos protótipos da ontologia.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 52 - Produtos requeridos na fase de manutenção

Da mesma forma como é feito na Methontology, esta proposta metodológica recomenda verificações e ajustes nas representações intermediárias (ver recomendações nas seções 6.2.4 e 6.2.5), caso ocorram modificações em elementos constitutivos das mesmas em função de alterações nos protótipos envolvidos no processo de desenvolvimento de ontologias. Desse modo, consegue-se evitar possíveis inconsistências em função de eventuais modificações.

A Tabela 53 mostra o resultado da fase de manutenção e o destino de tal resultado.

Informação de saída	Destino	
	Fase	Atividade
Recomendação de melhoria no processo de desenvolvimento da ontologia	Gerenciamento do projeto	Planejamento do protótipo
		Gerenciamento do protótipo

Tabela 53 - Produtos resultantes da fase de manutenção

6.2.8 Fase de Integração

Conforme discutido na fase de modelagem conceitual, especificamente na atividade de construção da estrutura taxonômica (seção 6.2.4.4), para viabilizar a construção de ontologias torna-se viável avaliar ontologias existentes, isto é, de alto nível e do próprio domínio, que forneçam termos pertinentes à conceitualização que está sendo construída.

A Tabela 54 relaciona os produtos requeridos ao processo de integração de forma que o mesmo seja realizado com eficiência.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Categorias ou facetas sintetizadas	Modelagem conceitual	Identificação de termos
		Definição de termos
		Agrupamento em facetas
Informação reportada sobre avaliação	Avaliação	Avaliação de resultados
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 54 - Produtos requeridos na fase de integração

Observa-se na Tabela 54 que a fase de integração recebe as categorias ou facetas sintetizadas resultantes da aplicação do método de análise facetada apresentado na seção 6.2.4.3. Tais facetas, conforme já discutido na seção 6.2.4.4, podem ser úteis no momento de avaliar as categorias de alto nível advindas de ontologias existentes. Desse modo, o mapeamento **de** facetas sintetizadas **para** categorias de alto nível pode ser avaliado através dos conceitos da estrutura facetada. Tal atividade pode ser auxiliada por documentos de integração, em que termos advindos da estrutura facetada são listados (coluna *termos na conceitualização construída* da Tabela 55) e relacionados com os termos advindos de ontologias avaliadas para reuso. Um exemplo de documento de integração é apresentado na Tabela 55.

Termos na conceitualização construída	Nome da ontologia avaliada para reuso: Ontolingua	
	Nome do termo	Descrição
Kilometer	Kilometer	Unidade padrão na Ontolingua
Centimeter	Indefinido	Unidade padrão na Ontolingua
Exponent	Expt	KIF-Números em Ontolingua

Tabela 55 – Modelo de um documento de integração
Fonte: Adaptado de Fernandez, Gomez-Perez e Juristo (1997, p.38)

No caso da extração de termos, pode-se optar pela integração automática, utilizando alguma ferramenta, como também pela extração manual de termos. Entretanto, é válido verificar até que

ponto os conceitos de outras ontologias podem ser integrados diretamente a ontologia de domínio. Almeida (2006, p.166) enfatiza que “o simples reaproveitamento de conceitos e relações de ontologias diferentes, a partir de uma análise superficial e automática, pode causar resultados inaceitáveis, visto que os escopos e contextos de criação das estruturas originais são, em geral, diferentes”.

Após a extração de termos e definições de outras ontologias terem sido realizados e registrados nos documentos de integração, cabe ao *ontologista* organizar os termos similares, provenientes das ontologias avaliadas, em conjunto com as suas definições. Deve-se escolher o termo e a definição mais adequada à conceitualização do domínio. Pode ser necessário ajustes como, por exemplo, tradução do termo para o idioma da conceitualização.

Uma vez que os termos e suas definições tenham sido determinados a partir do processo de integração, a estrutura taxonômica pode ser construída ou modificada (a modificação pode ocorrer em função da evolução de protótipos) em uma perspectiva *top-down*, *bottom-up* e *middle-up* de organização de conceitos, possibilitando, assim, que a conceitualização da ontologia possa representar o domínio de maneira consistente.

A Tabela 56 apresenta o produto resultante da fase de integração, que é destinado à fase de modelagem conceitual na atividade de construção da estrutura taxonômica.

Informação de saída	Destino	
	Fase	Atividade
Documento de integração	Modelagem conceitual	Construção da estrutura taxonômica
	Avaliação	Verificar pertinência

Tabela 56 - Produto resultante da fase de integração

6.2.9 Fase de Avaliação

O propósito da fase é fazer um julgamento técnico acerca dos produtos resultantes de cada fase durante todo o ciclo de vida da ontologia. A Tabela 57 relaciona tais produtos juntamente com suas fases e atividades que necessitam de um retorno, ou seja, após o procedimento de avaliação, cada fase recebe a informação do resultado da avaliação. Nesse caso, podem ser necessárias manutenções na estrutura da ontologia. Desse modo, cabe a fase de manutenção atuar em tal procedimento.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo
Cenários-problema	Pré-desenvolvimento	Análise da situação atual do domínio
Escopo da ontologia	Especificação de requisitos	Definição de escopo
Representações intermediárias	Modelagem conceitual	Conceitualização
Axiomas e regras formais	Formalização	Formalização de axiomas e regras
Ontologia implementada	Implementação	Implementação da ontologia
Documento de integração	Integração	Reuso de ontologias
Documentos gerados em todas as fases	Todas	Documentação
Modelo(s) para documentação	Documentação	Produção e distribuição de documentos
Documentação da fase	Documentação	Produção e distribuição de documentos

Tabela 57 - Produtos requeridos na fase de avaliação

Nessa proposta, dar-se-á ênfase à avaliação da ontologia implementada, que consiste em verificar tecnicamente a consistência da ontologia e assegurar que a mesma corresponda aos propósitos determinados. Para tal, o ambiente de desenvolvimento associado deve possibilitar efetuar consultas à ontologia através do mecanismo de inferência. Isso dá ao *ontologista* a capacidade de checar a consistência de sua ontologia, verificando, por exemplo, se ela responde às questões de competência relacionadas. Para que isso seja possível, é necessário que a linguagem de implementação da ontologia seja suportada pelo mecanismo de inferência adotado. Conforme foi visto na fundamentação teórica, a linguagem OWL está mapeada em lógicas descritivas e faz uso de inferências existentes como FaCT (*Fast Classification of Terminologies*) e RACER, possibilitando, desse modo, a utilização de inferências automáticas existentes em tais formalismos. O *Protégé-owl*[147] possui uma API, de *Application Programming Interface* (ou Interface de Programação de Aplicativos), que possibilita inferências sobre classes e instâncias na ontologia escrita em OWL.

Desta forma, através de consultas à ontologia, o *ontologista* pode verificar se a ontologia responde às questões de competência e, portanto, se satisfaz os requisitos para ela definidos.

A Tabela 58 apresenta de forma resumida o resultado da fase de avaliação que será submetido à fase e atividade requerente.

Informação de saída	Destino	
	Fase	Atividade
Resultados da avaliação	Todas	Específica de cada fase

Tabela 58 - Produto resultante da fase de avaliação

[147] http://protege.stanford.edu/plugins/owl/api/ReasonerAPIExamples.html. Acessado em 15 de maio de 2008.

6.2.10 Fase de Documentação

O propósito da fase é elaborar documentos de acordo com o planejamento da documentação advindo da fase de gerenciamento do projeto, conforme pode ser visualizado na Tabela 59.

Informação de entrada	Origem	
	Fase	Atividade
Plano de desenvolvimento	Gerenciamento do projeto	Planejamento do protótipo
Plano de gerenciamento	Gerenciamento do projeto	Gerenciamento do protótipo

Tabela 59 - Produtos requeridos na fase de documentação

Esta proposta metodológica compartilha da idéia da Methontology em incluir a documentação como uma atividade a ser realizada durante todo o processo de desenvolvimento da ontologia, a fim de se obter uma memória do projeto. Assim, para a fase de gerenciamento do projeto, vários documentos poderiam ser planejados para a gestão; para a fase pré-desenvolvimento, um documento seria direcionado aos cenários-problema; para a fase de especificação de requisitos seria criado um documento de especificação das questões de competência, de respostas às questões e de escopo da ontologia; para as fases de modelagem conceitual e formalização seriam criados documentos para o modelo conceitual e formal, incluindo o conjunto de representações intermediárias que descrevem a conceitualização do domínio; após a integração, um documento de integração; após a implementação, um documento de implementação da ontologia; após a manutenção, um documento sobre o que foi modificado; e após a avaliação, um documento sobre os resultados avaliados.

De acordo com o cenário exposto acima, modelos precisariam ser elaborados no sentido de serem distribuídos em cada fase do projeto para fins de preenchimento; posteriormente, os documentos resultantes de cada fase seriam organizados e arquivados para possíveis consultas. A Tabela 60 mostra os produtos resultantes da fase de documentação e seus destinos.

Salienta-se que esta proposta metodológica sugeriu uma série de modelos para a atividade de documentação, os quais foram baseados em trabalhos de autores de metodologias para construção de ontologias, bem como em experiências pessoais advindas da área de Engenharia de software. Finalmente, cabe ao gestor de projetos de ontologias determinar o formato de documentos mais adequado à sua realidade de trabalho.

Informação de saída	Destino	
	Fase	Atividade
Modelo(s) para documentação	Todas as fases	Produção e distribuição de documentos
Documentação	Todas as fases	Produção e distribuição de documentos

Tabela 60 - Produtos resultantes da fase de documentação

6.3 Considerações sobre a proposta metodológica para construção de ontologias

A proposta metodológica remete a reflexões acerca da questão interdisciplinar envolvida no contexto de construção de ontologias. A Tabela 61, apresentada no final dessa seção, mostra um panorama que permite visualizar as ponderações elucidadas a seguir.

O primeiro ponto a ponderar seria no modelo de ciclo de vida, que reflete o caminho que o desenvolvimento da ontologia deveria seguir e o modo como o desenvolvimento da ontologia é feito. Nesse sentido, a área de **Engenharia de Software** traz contribuições através dos modelos de processo de *software* propostos. A criação de um modelo de ciclo de vida ajuda a equipe de desenvolvimento a encontrar inconsistências, redundâncias e omissões no processo e em suas partes constituintes. À medida que esses problemas são percebidos e corrigidos, o processo se torna mais efetivo e concentrado na construção do produto final. Na proposta metodológica para construção de ontologias, o modelo de ciclo de vida adotado é o de *prototipação*, que permite modificações ao longo do processo e, com isso, a evolução da ontologia.

O segundo ponto pondera as técnicas de planejamento e de gestão de projeto advindas da **Engenharia de Software** que permitem construir produtos de qualidade, encontrar antecipadamente problemas, através de avaliações em cada fase do ciclo de vida, e seguir as restrições do orçamento e do cronograma. Na proposta metodológica foram recomendados estudo de viabilidade, análise de cenários e escalonamento de atividades envolvidas no processo de construção.

O terceiro ponto recai nos requisitos, que seriam a entrada do processo desenvolvimento e a saída em formato de uma ontologia. Para tal, a **Engenharia de Software** contribui mais uma vez com a técnica de análise de cenários, que permite a especificação das questões de competência, que por sua vez são propostas pela área de **Engenharia ontológica**, no sentido de verificar se a ontologia possui conhecimento suficiente para respondê-las.

O quarto ponto retrata os modelos conceituais criados para obter maior entendimento da entidade real a ser construída. A entidade, nesse caso, é a ontologia. Nesse sentido, o modelo deve ser capaz de representar o conhecimento da ontologia através de atividades de conceitualização, advindas da área de **Engenharia ontológica**, em que vários métodos e técnicas interdisciplinares são empregados para a construção da conceitualização. Dentre esses métodos e técnicas cabe destacar as contribuições da **Ciência da Informação** com a técnica de análise de assunto na extração de conceitos relevantes de documentos, refletindo, assim, em uma abordagem *bottom-up* na identificação de termos; a análise de conceito para realização de definições, que inicia pela observação do referente, seguindo para um levantamento de suas intensões e extensões para, finalmente, chegar à denominação através do termo; a organização dos termos através de facetas ou

categorias, que possibilitam o entendimento da natureza dos conceitos, viabilizando, desse modo, a construção da estrutura taxonômica da ontologia a partir de abordagens *middle-out* e *top-down* para organização de conceitos. A **Engenharia de Software** contribui com métodos utilizados na construção consensual do conhecimento como o método JAD, que possibilita discussões em grupo e decisões consensuais acerca da terminologia do domínio. Além disso, tal área contribui com padrões de modelagem bem aceitos na indústria, como é o caso da UML, que fornece construtos abstratos para modelagem conceitual de domínios através de representações gráficas.

O quinto ponto diz respeito às técnicas de representação do conhecimento oriundas da **Inteligência Artificial** e que dão condições à área de **Engenharia do Conhecimento** criar bases de conhecimento para uma dada finalidade. A base de conhecimento, nesse caso, seria representada por ontologias, que necessitam de formalismos lógicos a fim de serem especificadas. Tais formalismos, como lógica de primeira ordem, lógica descritiva, dentre outros, permitem a especificação de axiomas formais para restringir a interpretação dos fatos na ontologia.

O sexto ponto reflete a **Inteligência Artificial** e a **Engenharia ontológica** na fusão de fundamentos teóricos e metodológicos na construção de linguagens de implementação de ontologias no contexto da Web Semântica. As linguagens DAML+OIL e OWL possuem fundamentos em lógicas descritivas e objetivam definir, integrar e suportar ontologias de forma a proporcionar um entendimento comum sobre os conceitos usados nas páginas Web.

O sétimo ponto considera a evolução de protótipos no processo de construção de ontologias, o qual reflete em modificações ou manutenções na estrutura da ontologia em função de possíveis problemas identificados no processo de avaliação. Nesse sentido, a norma IEEE-1074 da **Engenharia de Software** propõe recomendar melhorias nas atividades de planejamento e de gerenciamento de projeto, a fim de melhorar a qualidade do produto.

O oitavo ponto analisa os três processos integrais, representados pela integração, avaliação e documentação, cujos métodos e técnicas refletem princípios teóricos da Ciência a Informação, bem como da Computação. O processo de *integração* parte de princípios da **Engenharia ontológica** em reutilizar conhecimento de ontologias de alto nível para representar um determinado domínio; utiliza-se de uma estrutura facetada, concebida a partir do método de análise facetada da **Ciência da Informação**, para auxiliar no mapeamento de categorias sintetizadas para categorias de alto nível advindas de ontologias existentes; e pode utilizar algoritmos especializados da área de **Inteligência Artificial** para realização de integração automática entre ontologias. O processo de *avaliação* segue princípios da **Engenharia de Software**, especificamente da norma IEEE-1074, em que avaliações devem ser realizadas ao final de cada atividade envolvida no processo de desenvolvimento. Isso possibilita que problemas sejam encontrados e corrigidos antes do início da fase de projeto. No

processo de verificação da consistência da ontologia, técnicas de **Inteligência Artificial**, como algoritmos de inferências, são empregadas nas consultas à ontologia para avaliar se a mesma responde às questões de competência, possibilitando assim, validá-la diante do escopo pré-definido. O processo de *documentação* utiliza de princípios da **Engenharia de Software**, baseados na norma IEEE-1074, para produzir documentação em cada fase do processo de desenvolvimento. Tal processo tem como propósito criar uma memória de projeto que se mostra útil em futuras manutenções e atualizações na ontologia.

Fases do ciclo de vida			Princípios teóricos e metodológicos da Ciência da Informação	Princípios teóricos e metodológicos da Ciência da Computação		
				Engenharia de software	Engenharia do conhecimento e IA	Engenharia ontológica
Gerenciamento do Projeto				Modelos de ciclo de vida. Planejamento e gestão de projetos.		
Processos orientados	Pré-desenvolvimento			Estudo de viabilidade. Análise de cenários		
	Desenvolvimento	Especificação de requisitos		Análise de cenários		Questões de competência
		Modelagem conceitual	Análise de assunto. Análise facetada com princípios da teoria da classificação e da teoria do conceito. Análise do conceito com princípios da teoria da terminologia e da teoria do conceito.	Método *Joint Application Design* – JAD. Linguagem de modelagem unificada (UML).	Técnicas de aquisição de conhecimento.	Atividades de conceitualização. Abordagem *top-down*, *middle-out* e *bottom-up*.
		Formalização			Técnicas de IA para representação do conhecimento	Especificação de axiomas e regras para restringir terminologias ou vocabulários
		Implementação			Técnicas de IA para representação do conhecimento	Linguagens de representação de ontologias
	Pós-desenvolvimento	Manutenção		Princípios da norma IEEE-1074		
Processos integrais	Integração		Análise facetada como auxílio à integração		Técnicas de IA para integração automática	Reuso de conhecimento
	Avaliação			Princípios da norma IEEE-1074	Técnicas de IA aplicadas nos mecanismos de inferência	Questões de competência
	Documentação			Princípios da norma IEEE-1074		

Tabela 61 – Panorama interdisciplinar da proposta metodológica para construção de ontologias

Finalmente, é válido destacar que a proposta metodológica para construção de ontologias cumpriu com os requisitos elucidados na seção 6.1. O **primeiro** requisito vislumbrava uma estratégia para a construção de ontologias fundamentada na norma internacional IEEE-1074 (1997). Tal estratégia foi determinada pelo modelo de ciclo de vida baseado em evolução de protótipos. O **segundo** requisito estava voltado à estratégia de construção de ontologias de domínio, que precisaria de uma conceitualização formada por conceitos de alto nível e conceitos comuns ao domínio. Tal requisito foi atendido pelas atividades de conceitualização propostas. O **terceiro** requisito vislumbrava um ciclo de vida que contemplasse o momento em que as atividades deveriam ser conduzidas; quais atividades seriam executadas em cada fase; e como estas fases estariam relacionadas. Tal requisito foi cumprido através da proposta de escalonamento de atividades em cada fase e pela explicitação do fluxo de informação de entrada e de saída nas fases e atividades envolvidas no processo de desenvolvimento. O **quarto** requisito demandava a explicitação dos princípios teóricos e metodológicos de cada campo do conhecimento envolvido na proposta. Tal requisito foi satisfeito através do panorama explicitado na Tabela 61.

7 – Conclusões e Perspectivas Futuras

Esta pesquisa de mestrado possibilitou percorrer diversos campos do conhecimento como a Filosofia, a Terminologia, a Ciência da Informação e a Ciência da Computação, no propósito de buscar contribuições acerca de tais campos para fornecer uma proposta de solução para problemas voltados à construção de ontologias. Tais problemas residem na falta de um padrão para construção destas e na falta de explicações sistemáticas de como, onde e sob quais limites podem ser utilizadas as abordagens teóricas dentro do processo de elaboração. Tal fato foi constatado na análise das metodologias e dos métodos para construção de ontologias investigados na pesquisa, que, na maioria dos casos, mostraram-se pouco eficientes na exposição clara dos procedimentos de construção. Desse modo, a solução para tais problemas estaria centrada em uma proposta metodológica fundamentada em princípios teóricos e metodológicos que dessem sustentação científica no processo de construção de ontologias.

A pesquisa traz uma inegável contribuição no sentido de propor princípios metodológicos para construção de ontologias fundamentados em teorias e metodologias advindas da Ciência da Informação, bem como da Ciência da Computação, respondendo, portanto, de forma afirmativa à questão de pesquisa "*Será possível confirmar a possibilidade de proposição de princípios metodológicos para construção de ontologias, à luz das contribuições pertinentes das Ciências da Informação e da Computação, dentre outros campos do conhecimento?*". Os resultados obtidos na pesquisa também possibilitaram comprovar os pressupostos elucidados na seção 1.1: o aporte teórico e metodológico existente no âmbito das Ciências da Informação e da Computação pode contribuir na proposição de metodologias para o processo de elaboração de ontologias, além das possibilidades de se obter contribuições advindas da Filosofia e da Terminologia. Desse modo, a pesquisa mostrou a correlação entre organização e uso da informação e técnicas computacionais no processo de construção de ontologias.

Ressalta-se que outros campos como a Filosofia e a Terminologia estão intrinsecamente ligados às Ciências da Informação e da Computação. A Filosofia, dentre outras questões, trata das leis de raciocínio, representadas pela lógica, que oferece declarações sobre todos os tipos de coisas no mundo e sobre as relações entre elas. Nesse sentido, contribui para o campo da Ciência da Computação, especificamente para a área de Inteligência Artificial na construção de algoritmos de inferência. O mesmo ocorre com os silogismos, que são contribuições importante do filósofo Aristóteles às áreas de Inteligência Artificial e a Engenharia ontológica no sentido de ser um padrão formal para representar regras de inferências e validar ontologias. A Terminologia antecipa as pesquisas sobre a análise de conceitos de Dahlberg em sua teoria do conceito, e propõe a sistematização de conceitos para a elaboração de definições consistentes sobre eles.

Ainda, como contribuições deste trabalho, têm-se:

- Considera-se que os princípios metodológicos propostos podem ser usados em qualquer domínio de conhecimento, pois as atividades de conceitualização envolvidas abrangem uma estrutura taxonômica com conceitos de alto nível e comuns a um domínio. Tal hipótese poderá ser verificada em trabalhos futuros a partir de uma implementação de uma ontologia de domínio.
- A definição de uma ontologia de domínio a partir dessa proposta possibilita que conceitos já estabelecidos possam ser reusados em domínios comuns.
- A criação de um arcabouço teórico-metodológico como um passo preliminar pode ser usado para a definição de novos padrões metodológicos para construção de ontologias.

Como extensão deste trabalho, as atividades pertencentes às fases da proposta poderiam ser mais bem especificadas, de forma a facilitar o processo de construção. São os casos das atividades voltadas à fase de gerenciamento do projeto, em que métodos e técnicas de Engenharia de Software, como métricas de produto e de processo, poderiam ser estudados no sentido de convergirem para metodologias de construção de ontologias. Na fase de modelagem conceitual poder-se-ia refinar a proposta da estrutura facetada para organização do domínio, explorando-se dos princípios e dos fundamentos da teoria da classificação de Ranganathan, bem como o campo da Lingüística, nos estudos da estrutura da linguagem. Desse modo, tornar-se-ia possível propor métodos mais consistentes para determinação de facetas e de seus indicadores, indispensáveis na construção de classificações analítico-sintéticas. Nas fases de formalização e implementação poder-se-ia conduzir um estudo sobre mecanismos de inferência, a fim de buscar conhecer os algoritmos mais eficientes para deduções a partir de seus formalismos. Desse modo, o processo de avaliação de ontologias poderia ser melhorado no sentido de verificar se os requisitos ou questões de competência estão sendo atendidos. Estudos sobre a criação de axiomas em linguagens de implementação de ontologias também poderiam ser conduzidos no sentido de selecionar a linguagem mais adequada no processo de construção.

Ainda, como trabalhos futuros, podem-se citar:

- Integrar a proposta metodológica a um editor de ontologias, de modo a permitir construir ontologias de domínio.
- Integrar a proposta metodológica a um ambiente de ferramentas para modelagem, de modo que a conceitualização possa ser visualizada de maneira gráfica.
- Aplicar a proposta metodológica a uma série de casos e necessidades reais, com o intuito de certificá-la e refiná-la.

Esses serão os objetivos a serem perseguidos de agora em diante.

Referências

ABADAL, E. CODINA Lluis. *Bases de datos documentales*: características, funciones y método. Madrid: Síntesis, 2005.

ALMEIDA, M.B. *Interoperabilidade entre fontes heterogêneas*: um meta-modelo baseado em ontologias. 149 f. Dissertação (Mestrado em Ciência da Informação) - Escola da Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2002.

ALMEIDA, M.B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, v. 32, n. 3, p.7-20, set./dez. 2003.

ALMEIDA, M.B. *Um modelo baseado em ontologias para representação da memória organizacional*. 2006. 341f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

ALVARENGA, Lídia. A teoria do conceito revisitada em conexões com ontologia e metadados no contexto das bibliotecas tradicionais e digitais. *Data Grama Zero*, v.2, n.6, dez. 2001. Artigo 05. Disponível em: <http://www.dgz.org.br>. Acesso em: 05 maio 2005.

ALVARENGA, Lídia. *Representação do conhecimento da perspectiva da Ciência da Informação em tempo e espaço digitais*. Enc. Bibli: R. Eletron. Bilbiotecon. Ci. Inf. Florianópolis, n.15, 2003. Disponível em: <www.encontros-bibli.ufsc.br/ Edicao_15/alvarenga_representacao.pdf>. Acesso em: 20 junho 2006.

ANSI Z39.19:1993. *American National Standard Guidelines for Thesuarus Structure, Construction, and Use*. New York, 1993.

ANSI/NISO Z 39.19: 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005. 176 p. Disponível em: <http://www.niso.org/standards/resources/Z39-19-005.pdf?CFID=2436046&CFTOKEN=81298864>. Acesso em: 27 jun. 2007.

ANTONIOU, G. OWL: *web ontology language*. In: ANTONIOU, G.; VAN HARMELEN, F. (Ed.). A semantic web primer. Cambridge: MIT, 2004. p. 109-150.

ARAUJO, V. R. M. H. *Sistemas de recuperação da informação:* nova abordagem teórico conceitual. 1994. 122f. Tese (Doutorado em Comunicação e Cultura) - Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1994.

BAADER, F. et al. *Terminological knowledge representation: a proposal for a terminological logic*. 1992. Disponível em: <http://citeseer.ist.psu.edu/baader92terminological.html>. Acesso em: 20 Fevereiro 2008.

BAADER, F., HORROCKS, I., SATTLER, U.: *Description logics as ontology languages for the semantic web*. In Staab, S., Studer, R., eds.: Lecture Notes in Artificial Intelligence. Springer Verlag, 2003. Disponível em: < http://citeseer.ist.psu.edu/baader03description.html>. Acesso em: 01 fevereiro 2008.

BARBOSA, Alice Príncipe. Classificações facetadas. *Ciência da Informação*, Rio de Janeiro, v. 1, n. 2, p. 73-81, 1972.

BARDIN, Laurence. *Análise de conteúdo*. Lisboa: Edições 70; 1977.

BARITE, Mario Guido. The Notion of “Category”: Its Implications in Subject Analysis and in the Construction and Evaluation of Indexing Language. *Knowledge Organization*, v. 27 (1/2), p.4-10, 2000.

BAX, Marcelo Peixoto. Introdução às linguagens de marcas. *Ciência da Informação*, Brasilía, v. 30, n.1, p. 32-38, jan./abr. 2001.

BECHHOFER, Sean; HORROCKS, Ian; PATEL-SCHNEIDER, Peter F. *Tutorial on OWL*. 2003. Disponível em: <http://www.cs.man.ac.uk/~horrocks/ISWC2003/Tutorial/>. Acesso em: 19 setembro 2007.

BENNER, K.M. et al. *Utilizings scenarios in the software development process*. 1993. Disponível em: <http://citeseer.ist.psu.edu/benner93utilizing.html>. Acesso em: 20 fevereiro 2008.

BERNARAS, A.; LARESGOITI, I.; CORERA, J. *Building and Reusing Ontologies for Electrical Network* Applications. In: PROCEEDINGS OF THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI/96, p. 298-302, 1996.

BERNERS-LEE, T; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, vol. 284, nº. 5, maio 2001, p. 34-43.

BERNERS-LEE, T; *Design Issues – Architectural and Philosophical Points*. 2003. Disponível em: <http://www.w3.org/DesignIssues/Stack.html>. Acesso em: 20 novembro 2007.

BERTALANFFY, Ludwig Von. *Teoria Geral dos Sistemas*. Petrópolis (RJ): Editora Vozes Ltda, 1968.

BOOCH, G. JACOBSON, I. RUMBAUGH, J. *UML Guia do Usuário*. Editora CAMPUS. Rio de Janeiro. 2006

BORST, W.N. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. 1997. Tese (Phd). Disponível em: <http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>. Acesso em: 03 outubro 2005.

BRAGA, M.; GUERRA, A.; REIS, J.C. *Breve história da ciência moderna*, vol.1, Jorge Zahar Editor, Rio de Janeiro, 2003.

BRAGANHOLO, Vanessa; HEUSER, Carlos. *XML Schema, RDF(S) e UML: uma comparação*. In: IDEAS 2001 - 4th Iberoamerican Workshop on Requirements Engineering and Software Environments , Santo Domingo, Heredia, Costa Rica, 2001.

BRANCHMAN R. J; SCHMOLZE J. G. An overview of the KL-ONE. *Journal of Knowledge Representation System.Cognitive Science*. v. 9. n. 2. p. 171-216. 1985.

BREITMAN, Karin. *Web semântica*: a internet do futuro. Rio de Janeiro: LTC, 2005.

BREU, Ruth; GROSU, Radu; HUBER, Franz; RUMPE, Bernhard; SCHWERIN, Wolfgang. *Towards a precise semantics for object-orientedmodeling techniques*. In Haim Kilov and Bernhard Rumpe, editors, Proceedings ECOOP’97 Workshop on Precise Semantics for Object- Oriented Modeling Techniques, pages 53–59. Technische Universit¨at M¨unchen, TUM-I9725, 1997.

BROCKMANS, S.; HAASE, P. *A Metamodel and UML Profile for Rule-extended OWL DL Ontologies {A Complete Reference. Technical report*, UniversitÄat Karl-sruhe, March 2006. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/sbr/publications/owl-metamodeling.pdf>. Acesso em: 05 Março 2008.

BUCKLAND, Michel. Information as Thing. *Journal of American Society of Information Science*, v.42, n.5, p. 351-360, 1991.

BUSH, Vannevar. As we may thing. *Atlantic Monthly*, v.176, n.1, p.101-108. 1945.

CALVANESE, D., GIACOMO, De G. *Descript ion Logics for Conceptual Data Modeling in UML.* ESSLLI, Vienna, August 18–22, 2003.
Disponível em: <http://www.dis.uniroma1.it/~degiacom/didattica/esslli03/>. Acesso em: 02 Abril 2008.

CAMPOS, M.L.A. *Linguagem Documentária*: Teorias que fundamentam sua elaboração. Niterói, RJ: EUFF, 2001.

CAMPOS, M.L.A. *A organização de unidades do conhecimento em hiperdocumentos*: o modelo conceitual como um espaço comunicacional para realização da autoria. 186p. Tese (Doutorado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Universidade Federal do Rio de Janeiro/ECO, Rio de Janeiro, 2001a.

CAMPOS, M.L.A.; BRASIL, M.I.; COELHO, B.A.S.; BASTOS, D. R.. *Vocabulário Sistematizado*: a experiência da Fundação Casa de Rui Barbosa. In: INTEGRAR, 2002, São Paulo. INTEGRAR. São Paulo, 2002.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 22-32, jan./abr. 2004.

CAMPOS, M.L.A.; GOMES, H.E.; MOTTA, D.F. *Tutorial de Tesauro*. 2004. Disponível em: http://www.conexaorio.com/biti/tesauro Acesso em: 30 abril 2008.

CAMPOS, M.L.A.; GOMES, H.E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 3, p.348-359, set./dec. 2006.

CAPURRO, R. *Epistemologia Y ciencia de la información*. IN: Encontro Nacional de Pesquisa em Ciência da Informação, 5, 2003, Belo Horizonte. Anais. Belo Horizonte, Escola de Ciência da Informação/UFMG.

CARDOSO, J. The Semantic Web Vision: Where are We? *IEEE Intelligent Systems*, p. 22-26, Set.Out. 2007.

CERQUEIRA, Roberto Figueiredo Paletta de. *Método de modelagem domínio-ontológica do Direito Positivo Brasileiro*. 2007. 131f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2007.

CHANDRASEKARAN, B.; JOHNSON, T. R.; BENJAMINS, V. R. Ontologies: what are they? why do we need them?. *IEEE Intelligent Systems*, Washington, v. 14, n. 1, p. 20-26, Jan./Fev.1999.

CHEN, Peter P. *The Entity-Relationship Model - Toward a Unified View of Data*. 1976. Disponível em: <http://www.csc.lsu.edu/news/erd.pdf>. Acesso em: 24 Maio 2007.

CINTRA, A. M. M. ; KOBASHI, N. Y. ; LARA, M. L. G. ; TÁLAMO, M. F. M. G. *Para entender as linguagens documentárias*. 2ª ed. São Paulo: Editora Pólis, 2002.

CODD, E.F. *A Relational Model of Data for Large Shared Data Banks*. Communications of the ACM 13 (6): 377–387. 1970.
Disponível em <http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf >. Acesso em: 02 Abril 2007.

COPI, I.M. *Introdução a lógica*. 2ed. São Paulo: Mestre Jou, 1981.488p.

CORAZZON, R. *What is Ontology? Definitions by leading philosophers*. In: Ontology. A Resource Guide for Philosophers. 2008. Disponível em <http://www.formalontology.it/section_4.htm>. Acesso em: 02 Abril 2008.

CORAZZON, R. *The Theory of Predication in Aristotle's Categories*. In: Ontology. A Resource Guide for Philosophers. 2008a. Disponível em: <http://www.formalontology.it/aristotle-categories.htm>. Acesso em: 10 Abril 2008.

CORCHO, O.; GÓMEZ-PÉREZ, A.; LÓPEZ-CIMA, A.; LÓPEZ-GARCÍA, V.; SUÁREZ-FIGUEROA, M. C.: ODESeW. *Automatic Generation of Knowledge Portals for Intranets and Extranets*. 2003. Disponível em: <http://webode.dia.fi.upm.es/sew/>. Acesso em: 05 abril 2008.

CORNELIUS, Ian.Theorizing information for Information science. *Annual Review of Information science and Technology*.v.36. 2002. P.393-425.

CRANEFIELD, S.; PURVIS, M. *UML as an ontology modeling language*. In: WORKSHOP ON INTELLIGENT INFORMATION INTEGRATION, 16th, 1999, Stockholm. Proceedings… Stockholm: International Joint Conference on Artificial Intelligence (IJCAI-99), 1999. p.1-16.

CURRÁS, Emília. *Tesauros*: linguagens terminológicas. Brasília, IBICT, 1995. 286 p.

CYC Project. *What's in Cyc?* 2005. Disponível em: <http://www.cyc.com/cyc/technology/whatiscyc_dir/whatsincyc>. Acesso em: 20 abril 2008.

DAHLBERG, Ingetraut. Classification theory, yesterday and today. *International classification*, v. 3, n. 2, p. 85-90, 1976.

DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v. 7, n. 2, p. 101-107, jul./dez. 1978.

DEAN M., CONNOLLY D., VAN HARMELEN F., HENDLER J., HORROCKS I., MCGUINNESS D.L., PATEL-SCHNEIDER P.F., and STEIN L.A. *OWL Web Ontology Language 1.0 Reference*. W3C Working Draft 21. 2003.

DODEBEI, Vera Lucia Doyle. *Tesauro*: linguagem de representação da memória documentária. Niterói; Rio de Janeiro: Intertexto; Ed. Interciência, 2002.

DONINI, F.M.; LENZERINI, M.; NARDI, D.; NUTT, W. *The complexity of concept languages, Information and Computation*, 134(1), 1997.
Disponível em: http://citeseer.ist.psu.edu/donini91complexity.html. Acesso em: 03 abril 2008.

EVANS, A.. 1998. *Reasoning with UML diagrams*. Proceedings of the Workshop on Industrial Strength Formal Methods (WIFT'98). IEEE Press.

EVANS, Andy; FRANCE, Robert; LANO, Kevin; RUMPE, Bernhard. *Developing the UML as a formal modelling notation*. In Pierre-Alain Muller and Jean B´ezivin, editors, Proceedings of UML'98 International Workshop, Mulhouse, France, June 3 - 4, 1998, pages 297–307. ESSAIM, Mulhouse, France, 1998.

FALBO, Ricardo de Almeida. *Integração de conhecimento em um ambiente de desenvolvimento de software*. 81 f. Tese (Doutorado em Ciências) - COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.

FENSEL, D. et al. OIL: an ontology infraestructure for the semantic web. *IEEE Intelligent Systems*, Washington, v. 16, n. 2, p. 38-44, Mar./April 2001.

FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, H. *Methontology:* from ontological art towards ontological engineering. 1997.
Disponível em: <http://citeseer.ist.psu.edu/context/544607/0/>. Acesso em: 20 Agosto 2007.

FERNÁNDEZ, M; GOMEZ-PEREZ, A.; SIERRA, J.P.; SIERRA, A.P . Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, jan./ fev.1999.

FERNÁNDEZ, M. *Overview of methodologies for building ontologies*. 1999. Disponível em: < http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>. Acesso em: 16 Novembro 2007.

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

FISCHER, D., 1998. *From Thesauri towards Ontologies?* In W. M. el Hadi, J. Maniez, and St. A. Pollit (eds.), Structures and Relations in Knowledge Organization. Proceedings of the 5th ISKO-Conference, Lille. W¨urzburg: Ergon Verlag.

FOX, M.S. *The TOVE Project:* towards a common-sense model of the enterprise. 1992. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/index.html>. Acesso em: 20 Novembro 2007.

FOX, M.S. et al. *An organization ontology for enterprise modelling*. 1997. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/org-prietula-23aug97.pdf>. Acesso em: 10 Abril 2008.

FREIRE, Gustavo Henrique de A. Ciência da Informação: temática, histórias e fundamentos. *Perspectivas em Ciência da Informação*, Vol. 11, N° 1, 2006. Disponível em: <http://www.eci.ufmg.br/pcionline/viewarticle.php?id=406>. Acesso em: 02 Março 2008.

FREITAS, Frederico L. G. *Ontologias e a Web Semântica*. Santos: Universidade Católica de Santos, 2003. Disponível em: <http://www.inf.unisinos.br/~renata/cursos/topicosv/ontologias-ws.pdf >. Acesso em: 22 Março 2008.

FUJITA, M.S.L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Belo Horizonte, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: < http://server01.bc.unicamp.br/seer/ojs/ viewarticle .php?id=9>. Acesso em: 03 Maio 2005.

FURLAN, José Davi. *Modelagem de objetos através da UML*: the unified modeling language. São Paulo: Makron Books, 1998.

GILCHRIST, Alan. Thesauri, taxonomies and ontologies - an etymological note. *Journal of Documentation*, v. 59, n. 1, 2003, p. 7-18.

GILLILAND-SWETLANDO, Anne J. *Introduction to metadata*: setting the stage. 2000. Disponível em: <www.getty.edu/research/conducting_research/ standards/intrometadata/pdf/swetland.pdf >. Acesso em: 12 Maio 2005.

GOMES, H.E. et al. *Manual de Elaboração de Tesauros Monolíngues*. Brasília: CNPq/PNBU, 1990. 78p.

GOMES, H.E. *Classificação, Tesauro e Terminologia*: fundamentos comuns. Palestra preparada para as Tertúlias do Departamento de Biblioteconomia da UNIRIO. Julho, 1996.

GÓMEZ-PÉRÉZ, A. *Some ideas and examples to evaluate ontologias*. 1994. Disponível em: <http://citeseer.ist.psu.edu/gomez-perez94some.html>. Acesso em: 12 Setembro 2007.

GÓMEZ-PÉREZ, A., FERNÁNDEZ, M., VICENTE, A.J. *Towards a Method to Conceptualize Domain Ontologies*, ECAI'96 - Workshop on Ontological Engineering, Budapest, August, 1996. Disponível em: <http://citeseer.ist.psu.edu/483876.html>. Acesso em: 12 Novembro 2007.

GÓMEZ-PÉREZ, A. *Evaluation of taxonomic knowledge in ontologies and knowledge bases*. In: WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, 1999, Alberta, Proccedings… Alberta: University of Calgary, 1999. p.6.1.1-6.1.18.

GRUBER, T.. *What is an Ontology?* 1993. Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. Acesso em: 03 Agosto 2006.

GRUBER,T. *A Translation Approach to Portable Ontology Specifications*. 1993a. Disponível em: <ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-92-71.ps.gz>. Acesso em: 03 Agosto 2006.

GRUBER, T. R. *Towards principles for the design of ontologies used for knowledge sharing*. In N. Guarino and R. Poli (Eds.), Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer. 1994. Disponível em: <http://citeseer.ist.psu.edu/gruber93toward.html>. Acesso em: 10 Março 2007.

GRUNINGER, Michael; FOX, Mark S. *The Design and Evaluation of Ontologies for Enterprise Engineering*. 1994. Disponível em: <http://citeseer.ist.psu.edu/353050.html>. Acesso em: 03 Novembro 2007.

GRUNINGER, Michael; FOX, Mark S. *The Role of Competency Questions in Enterprise Engineering*. 1994a. Disponível em: <http://citeseer.ist.psu.edu/6130.html>. Acesso em: 03 Novembro 2007.

GRUNINGER, M.; FOX, M. S. *Methodology for the design and evaluation of ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/grninger95methodology.html>. Acesso em: 10 Novembro 2007.

GRUNINGER, M.; FOX, M. S. *The Logic of Enterprise Modelling*. 1996. Disponível em: < http://citeseer.ist.psu.edu/523974.html>. Acesso em: 10 Novembro 2007.

GRUNINGER, M.; FOX, M.S. *Enterprise modeling*. 1998. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/>. Acesso em: 21 Novembro 2007.

GUARINO, N. *Formal ontology, conceptual analysis and knowledge representation*. 1995. Disponível em: <http://citeseer.ist.psu.edu/guarino95formal.html>. Acesso em: 03 Setembro. 2007.

GUARINO, N.; GIARETTA, P. *Ontologies and KBs, towards a terminological clarification*. In: MARS, N. (Ed.). Towards a Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing. [S.l.]: IOS Press, 1995. p. 25-32.

GUARINO, N. *Formal ontology in information systems*. 1998. Disponível em: <http://citeseer.ist.psu.edu/guarino98formal.html>. Acesso em: 03 Setembro 2007.

GUARINO, N.; WELTY, C. *Ontological Analysis of Taxonomic Relationships*. 2000. Disponível em: <http://citeseer.ist.psu.edu/guarino00ontological.html>. Acesso em: 12 setembro 2007.

GUINCHAT, C., MENOU, M. *As linguagens documentais*. In: ___. Introdução geral às ciências e técnicas da informação e documentação. Brasília: IBICT, 1994. p. 133-165.

HEFLIN, J.; HENDLER, J. *Searching the Web with SHOE*. 2000. Disponível em: <http://citeseer.ist.psu.edu/heflin00searching.html>. Acesso em: 20 Novembro 2007.

HORRIDGE, M. et al. *A Pratical Guide to Building OWL Ontologies Using Teh Protégé-OWL Plugin and CO-ODE Tools*. Ed.1. The Univerrity of Manchester and Stanford University, Agosto, 2004. Disponível em: <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>. Acesso em: 07 Fevereiro 2007.

HORROCKS, I. et al. *Reference description of the DAML+OIL ontology markup language*. 2001. Disponível em: <http://www.w3.org/TR/daml+oil-reference>. Acesso em: 04 Novembro 2007.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS: *IEEE Standard 1074 Standard for developing software life cycle processes*. 1997. Disponível em: < http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel4/5984/16018/00741936.pdf?temp=x>. Acesso em: 20 Julho 2007.

JACOB, E., K. Classification and categorization: a difference that makes a difference. *Library Trends/Winter*, 2004.

JIMENEZ, Antonio García. Instrumentos de Representación del Conocimiento: Tesauros versus Ontologías. *Anales de Documentación*, nº 7, p. 79-95, 2004.

JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development*. 1998. Disponível em: <http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf>. Acesso em: 20 Novembro 2007.

JONES, D.M.; PATON, R.C. *Toward Principles for the Representation of Hierarchical Knowledge in Formal Ontologies*. 1999. Disponível em: <http://citeseer.ist.psu.edu/306693.html> . Acesso em: 12 Novembro 2007.

JORGE, Marcelo Alvim. *Ontologias no suporte a portais semânticos*. Dissertação (Mestrado em Ciências da Informação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

JURISICA, I.; MYLOPOULOS, J.; YU, E. *Using Ontologies for Knowledge Management*: An Information Systems Perspective. Annual Conference of the American Society for Information Science, Washington, D.C.,1999.

KARP, R.; CHAUDHRI, V.; THOMERE, J. XOL. *an XML-based ontology exchange language*. 1999. Disponível em: <http://www.ai.sri.com/~pkarp/xol/xol.html>. Acesso em: 22 Novembro 2007.

KIFER, M.; LAUSEN, G.; WU, J. *Logical Foundations of Object-Oriented and Frame-Based Languages*. 1995. Disponível em: < http://www.cs.umbc.edu/771/papers/flogic.pdf >. Acesso em: 22 Março 2008.

KIM, H.M.; FOX, M.S.; GRUNINGER, M. *An ontology for quality management; enabling quality problem identification and tracing*. 1999. Disponível em: <http://citeseer.ist.psu.edu/context/1728409/0>. Acesso em: 20 Novembro 2007.

LANCASTER, F. W. *Vocabulary control for information retrieval*. 2. ed. Virgínia : IRP, 1986. 270 p.

LANCASTER, F. W.; WARNER, Amy J. Information retrieval today. Revised,retitled and expanded ed. Arlington: *Information Resources Press*, 1993.

LANCASTER, F.W. *Indexação e resumos*: teoria e prática. Brasília: Briquet de Lemos, 1993.

LASSILA, O.; SWICK, R. *Resource Description Framework* (RDF): model and syntax specification, W3C Recommendation. 1999. Disponível em: <http://www.w3.org/TR/REC-rdf-syntax/>. Acesso em: 25 Setembro 2007.

LAWRENCE, Steve S., BOLLACKER K., LEE GILES, C. Digital Libraries and Autonomous Citation Indexing. *IEEE Computer*, v. 32, n. 6, p.67-71, 1999.

LENAT, D. CYC: *A large-scale investment in knowledge infrastructure*. In: Communications of the ACM, 38(11):33-38. 1995. Disponível em: < http://citeseer.ist.psu.edu/lenat95cyc.html >. Acesso em: 10 Abril 2008.

LIMA, G. Â. B. O. *Mapa Hipertextual (MHTX)*: um modelo conceitual para organização hipertextual de documentos. 2004. 199 f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

LOCKHEED, Martin. 2000. *UBOT details*.
Disponível em: <http://ubot.lockheedmartin.com/ubot/details/uml_to_daml.html> . Acesso em: 10 Novembro 2007.

LOZANO-TELLO A. and GÓMEZ-PÉREZ A. ONTOMETRIC: A Method to Choose the Appropriate Ontology. *Journal Of Database Management*. 2004. Disponível em: <http://www.accessmylibrary.com/coms2/summary_0286-20574535_ITM>. Acesso em: 07 Novembro 2007.

MCBRIDE, B. The Resource Description Framework (RDF) and its Vocabulary Description Language RDFS. In: STAAB, S.; STUDER, R. (ED.). *Handbook on Ontologies*. Berlin: Springer-Verlag, 2004. p.51-66

MILLER, G.A. *WordNet: a lexical database for English*. 1995. Disponível em: <http://portal.acm.org/citation.cfm?id=219748>. Acesso em: 20 Março 2008.

MIZOGUCHI, R.; VANWELKENHUYSEN, J., IKEDA, M. *Task ontology for reuse of problem solving knowledge*. In: ECAI'94 Towards very large knowledge bases, 1995, Amsterdam, Procceddings… Amsterdam: IOS Press, 1995. p. 46-59

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A.P. Thesaurus and ontologies: a study over the definitions found in the computer and information science literature, by means of Analytical-synthetic Method. *Knowledge Organization*, v. 31, n. 4, p.231-244, 2004.

NARDON, Fabiane Bizinella. *Compartilhamento de Conhecimento em Saúde utilizando Ontologias e Bancos de Dados Dedutíveis*. 2003. 222f. Tese (Doutorado em Engenharia Elétrica) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2003.

NAVATHE, Shamkant B.; ELMASRI, Ramez. *Sistemas de bancos de dados*: fundamentos e aplicações. Rio de Janeiro: LTC, 2000.

NOVAK, J. D.; CAÑAS, A. J. *The Theory Underlying Concept Maps and How to Construct Them.* Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008.
Disponível em:
<http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>.
Acesso em: 23 Março 2008.

NOVELINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. *Informação & Informação*, Londrina, v. 1, n. 2, p. 37-45, jul./dez. 1996.

NOY, F. N.; GUINNESS, D. L. *Ontology development 101*: a guide to create your first ontology. 2001. Disponível em: <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>. Acesso em: 03 Outubro 2006.

OLIVEIRA, M. (Coord).*Ciência da Informação e Biblioteconomia*: novos conteúdos e espaços de atuação. Belo Horizonte: Ed.UFMG, 2005. 143p.

OMG. 2003. *Object Management Group*. www.omg.org

OVERGAARD, Gunnar. *A formal approach to relationships in the Unified Modeling Language*. In Manfred Broy, Derek Coleman, Tom S. E. Maibaum, and Bernhard Rumpe, editors, Proceedings PSMT'98 Workshop on Precise Semantics for Modeling Techniques. Technische Universit¨at M¨unchen, TUM-I9803, 1998.

PAULA FILHO, Wilson de Pádua. *Engenharia de software*: fundamentos, métodos e padrões. 2. ed. Rio de Janeiro: LTC, 2003. 602p.

PINTO, Paulo Roberto Margutti. *Apostila de lógica II*. Faculdade de Filosofia e Ciências Humanas. Departamento de Filosofia. Belo Horizonte, 2006. 113 p. Disponível em <http://www.fafich.ufmg.br/~margutti/Textos.html>. Acesso em: 02 fevereiro 2008.

PRESSMAN, Roger S**.** *Engenharia de Software*. 5.ed. Rio de Janeiro: McGraw-Hill, 2002. 843 p.

PRIETO-DÍAZ, R. *A Faceted Approach to Building Ontologies*, In Proceedings of the 2003 IEEE International Conference on Information Reuse and Integration, IRI'03, IEEE SMC(Las Vegas, NV), 27 Oct 2003.

RANGANATHAN, S.R. *Prolegomena to library classification*. Bombay: Asia Publishing House, 1967. 640p.

RANGANATHAN, S. R. *Facet analysis*: fundamental categories. In: CHAN, L. M., RICHMOND, P., SVENONIUS, E. (org.) Theory of subject analysis: a sourcebook. Littleton, Co.: Libraries Unlimited, 1985. p. 86-93

REED, S.L.; LENAT, D.B. *Mapping Ontologies into Cyc*. 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 20 Abril 2008.

REZENDE, Solange Oliveira. *Sistemas Inteligentes*: fundamento e aplicações. São Paulo: Manole, 2003.

RICHTERS, Mark; GOGOLLA, Martin. *On formalizing the UML Object Constraint Language OCL*. In TokWang Ling, Sudha Ram, and Mong Li Lee, editors, Proc. 17th Int. Conf. Conceptual Modeling (ER'98). Lecture Notes in Computer Science, number 1507, Springer-Verlag, 1998.

RUMBAUGH, J., BLAHA, M. *Modelagem e projetos baseados em objetos com UML* 2. Rio de Janeiro : Campus, 2006.

RUSSELL, S., NORVIG, P. *Inteligência artificial*. 2. ed. Rio de Janeiro: Campus, 2004.

SARACEVIC, T. Ciência da Informação: origem, evolução, relações. *Perspectivas em Ciência da informação*, Belo Horizonte, v.1, n.1, p.41-62, jan/jun 1996.

SATIJA, M. P. – Library classification: an essay in terminology. *Knowledge Organization*. Vol. 27, n.4, p. 221-229, 2000.

SAYAO, L. F. Modelos teóricos em Ciência da Informação: abstração e método científico. *Ciência da Informação*, Brasília: UnB, v. 30, n. 1, p. 82-91, jan./abr. 2001

SCHREIBER, G.; WIELINGA, B; JANSWEIJER, W. *The Kactus View on the 'O' Word.* Workshop on Basic Ontological Issues in Knowledge Sharing/IJCAI95; Montreal, Canada; August 1995. Disponível em: < http://www.cs.vu.nl/~guus/papers/Schreiber95a.pdf > Acesso em: 15 Abril 2008.

SILBERSCHATZ, A., KORTH, H.F., SUDARSHAN, S. *Sistemas de Bancos de Dados*. Tradução: Daniel Vieira. 5ª Edição. Rio de Janeiro: Elsevier. 2006.

SMITH, B. *Ontology and Informations Systems*. 2004. Disponível em: <http://www.ontology.buffalo.edu/ontology> Acesso em: 15 Setembro 2007.

SOERGEL, Dagobert. *Functions of a Thesaurus / Classification / Ontological Knowledge Base*. College of Library and Information Services, University of Maryland. 1997.

SOERGEL, Dagobert. The Rise of Ontologies or the Reinvention of Classification. *Journal of the American Society of Information Science* 50(12): 1119-1120. 1999.

SOMMERVILLE, Ian. *Engenharia de Software*. 8a. edição. 2007. 592p.

SOUZA, Renato Rocha. *Uma proposta de metodologia para escolha automática de descritores utilizando Sintagmas Nominais*. 202p. Tese (Doutorado em Ciência da Informação) – Escola de Ciência daInformação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

SOUZA, Renato Rocha. Sistemas de Recuperação de Informações e Mecanismos de Busca na Web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 2, p.161-173, mai./ago. 2006.

SOWA, J.F. *Building, sharing and merging ontologies*. 1999. Disponível em: <http://users.bestweb.net/~sowa/ontology/ontoshar.htm>. Acesso em: 03 Outubro 2007.

SOWA, J. F. *Ontoloy*. 2000. Disponível em: <http://www.jfsowa.com/ontology/>. Acesso em: 20 Março 2008.

STEVENS, R. et al. Ontologies in Bioinformatics. In: STAAB, S.; STUDER, R. (Ed.). *Handbook on Ontologies*. Berlin: Springer-Verlag, p. 635-658. 2004.

SWARTOUT, B; PATIL, R.; KNIGHT, K.; RUSS, T. *Toward Distributed Use of Large-Scale Ontologies*. 1996. Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em: 10 Abril 2008.

TOLENTINO, Ricardo. *Aplicações WEB em XML*: Estágio Atual e tendências futuras. 2004.

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973. 37 p.

USCHOLD, M.; KING, M. *Towards a Methodology for Building Ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/uschold95toward.html> Acesso em: 10 Novembro 2007.

USCHOLD, M. *Building Ontologies:* Towards a Unified Methodology. 1996. Disponível em: < http://citeseer.ist.psu.edu/uschold96building.html > Acesso em: 10 Novembro 2007.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.

USCHOLD, M. et al. *Enterprise ontology*. 1998. Disponível em: <http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>. Acesso em: 10 Novembro 2007.

VALENTE, A.; RUSS, T.; McGREGOR, R.; SWARTOUT, W. Building and (Re) Using an Ontology of Air Campaign Planning. *IEEE Intelligent Systems.* Jan/Fev. 1999. Disponível em: < http://portal.acm.org/citation.cfm?id=630305.630437&coll=GUIDE&dl=GUIDE>. Acesso em: 12 Abril 2008.

VAN HEIJST, G.; SCHREIBER, A.T.; WIELINGA, B.J. *Using Explicit Ontologies in KBS Development*. 2002. Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/borst/node16.html >. Acesso em: 23 Abril 2007.

VASCONCELLOS, M.J.E. de. *Pensamento sistêmico*: o novo paradigma da ciência. Campinas: Papirus, 2002. 268p.

VELLUCI, Sherry L. Metadata. *Annual Review of Information Science and Technology* (*ARIST*), v.33, p.189-222, 1998.

VICKERY, B.C. *Classificação e indexação nas ciências*. Rio de Janeiro: BNG/Brasilart, 1980. p. 187-224: Aspectos históricos da classificação da ciência.

VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997.

WIELINGA, B. J.; SCHREIBER, A. T.; BREUKER, J.A. KADS*:* A modelling approach to knowledge engineering. *Knowledge Acquisition* (4). p. 5-53. 1992.

ZILLES, Urbano. *Teoria do Conhecimento*, Coleção Filosófica 21, 3ª. ed. Porto Alegre: EDIPUCRS, 167 pág., 1998.

Anexo A: Referências da amostra de documentos selecionados para análise – em ordem cronológica de publicação

R1 FOX, M.S. *The TOVE Project; towards a common-sense model of the enterprise*. 1992. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/index.html>. Acesso em: 20 Novembro 2007.

R2 GRUNINGER, Michael; FOX, Mark S. *The Design and Evaluation of Ontologies for Enterprise Engineering*. 1994. Disponível em: <http://citeseer.ist.psu.edu/353050.html>. Acesso em: 03 Novembro 2007.

R3 GRUNINGER, M.; FOX, M. S. *Methodology for the design and evaluation of ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/grninger95methodology.html>. Acesso em: 10 Novembro 2007.

R4 USCHOLD, M.; KING, M. *Towards a Methodology for Building Ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/uschold95toward.html> Acesso em: 10 Novembro 2007.

R5 BOUAUD, B.; BACHIMONT, B.; CHARLET, J.; ZWEIGENBAUM, P. *Methodological Principles for Structuring an "Ontology*". 1995. Disponível em: < http://citeseer.ist.psu.edu/bouaud95methodological.html> Acesso em: 10 Novembro 2007.

R6 GÓMEZ-PÉREZ, A., FERNÁNDEZ, M., VICENTE, A.J. *Towards a Method to Conceptualize Domain Ontologies*, ECAI'96 - Workshop on Ontological Engineering, Budapest, August, 1996. Disponível em: <http://citeseer.ist.psu.edu/483876.html>. Acesso em: 12 Novembro 2007.

R7 USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.

R8 USCHOLD, M. *Building Ontologies: Towards a Unified Methodology*. 1996. Disponível em: < http://citeseer.ist.psu.edu/uschold96building.html > Acesso em: 10 Novembro 2007.

R9 GRUNINGER, M.; FOX, M. S. *The Logic of Enterprise Modelling*. 1996. Disponível em: < http://citeseer.ist.psu.edu/523974.html>. Acesso em: 10 Novembro 2007.

R10 FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, H. *Methontology:* from ontological art towards ontological engineering. 1997.
Disponível em: <http://citeseer.ist.psu.edu/context/544607/0/>. Acesso em: 20 Novembro 2007.

R11 GRUNINGER, M.; FOX, M. S. *On Ontologies and Enterprise Modelling*. 1997. Disponível em: < http://citeseer.ist.psu.edu/fox97ontologies.html >. Acesso em: 10 Novembro 2007.

R12 CAROLE, H.; NOY, D.; FRIDMAN, N. *The state of the art in ontology design: a survey and comparative review*. 1997. Disponível em: < http://www.encyclopedia.com/doc/1G1-20418259.html>. Acesso em: 10 Novembro 2007.

R13 USCHOLD, M. et al. *Enterprise ontology.* 1998. Disponível em: <http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>. Acesso em: 10 Novembro 2007.

R14 JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development*. 1998. Disponível em: <http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf>. Acesso em: 20 Novembro 2007.

R15 FERNÁNDEZ, M. *Overview of methodologies for building ontologies*. 1999. Disponível em: < http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>. Acesso em: 16 Novembro 2007.

R16 JONES, D.M.; PATON, R.C. *Toward Principles for the Representation of Hierarchical Knowledge in Formal Ontologies*. 1999. Disponível em: <http://citeseer.ist.psu.edu/306693.html> . Acesso em: 12 Novembro 2007.

R17 KALFOGLOU, Y. *Deploying Ontologies in Software Design*. 2000. Disponível em: < http://citeseer.ist.psu.edu/kalfoglou00deploying.html > . Acesso em: 12 Novembro 2007.

R18 SURE, Y.; ANGELE, J.; STAAB, S. *OntoEdit: Guiding Ontology Development by Methodology And Inferencing*. 2002. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2002_odbase_ontoedit.pdf > . Acesso em: 12 Novembro 2007.

R19 SURE, Y.; STAAB, S.; STUDER, R. *Methodology for Development and Employment of Ontology based Knowledge Management Applications*. 2002. Disponível em: < http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2002_sigmod-methodology.pdf>. Acesso em: 12 Novembro 2007.

R20 SURE, Y.; STUDER, R. *On-To-Knowledge Methodology Final Version*. 2002. Disponível em: < http://citeseer.ist.psu.edu/540438.html>. Acesso em: 12 Novembro 2007.

R21 STEVENS, R.; GOBLE, C.; HORROCKS, I.; BECHHOFER, S. *Building a Bioinformatics Ontology Using OIL*. 2002.
Disponível em: < http://ieeexplore.ieee.org/iel5/4233/21713/01006301.pdf >. Acesso em: 12 Novembro 2007.

R22 SURE, Y.; STAAB, S.; STUDER, R. *On-To-Knowledge Methodology (OTKM)*. 2003. Disponível em: < http://citeseer.ist.psu.edu/578235.html>. Acesso em: 12 Novembro 2007.

R23 FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

R24 BREITMAN, Karin. *Web semântica*: a internet do futuro. Rio de Janeiro: LTC, 2005.

R25 CARDOSO, J. The Semantic Web Vision: Where are We? *IEEE Intelligent Systems*, p. 22-26, Set.Out. 2007.

Anexo B: Referências da literatura sobre o método Cyc

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

LENAT, D. CYC: *A large-scale investment in knowledge infrastructure*. In: Communications of the ACM, 38(11):33-38. 1995. Disponível em: < http://citeseer.ist.psu.edu/lenat95cyc.html >. Acesso em: 10 Abril 2008.

REED, S.L.; LENAT, D.B. *Mapping Ontologies into Cyc*. 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 20 Abril 2008.

Anexo C: Referências da literatura sobre a metodologia de Gruninger e Fox

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

FOX, M.S. *The TOVE Project:* towards a common-sense model of the enterprise. 1992. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/index.html>. Acesso em: 20 Novembro 2007.

FOX, M.S. et al. *An organization ontology for enterprise modelling*. 1997. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/org-prietula-23aug97.pdf>. Acesso em: 10 Abril 2008.

GRUNINGER, Michael; FOX, Mark S. *The Design and Evaluation of Ontologies for Enterprise Engineering*. 1994. Disponível em: <http://citeseer.ist.psu.edu/353050.html>. Acesso em: 03 Novembro 2007.

GRUNINGER, Michael; FOX, Mark S. *The Role of Competency Questions in Enterprise Engineering*. 1994a. Disponível em: <http://citeseer.ist.psu.edu/6130.html>. Acesso em: 03 Novembro 2007.

GRUNINGER, M.; FOX, M. S. *Methodology for the design and evaluation of ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/grninger95methodology.html>. Acesso em: 10 Novembro 2007.

GRUNINGER, M.; FOX, M. S. *The Logic of Enterprise Modelling*. 1996. Disponível em: < http://citeseer.ist.psu.edu/523974.html>. Acesso em: 10 Novembro 2007.

GRUNINGER, M.; FOX, M.S. *Enterprise modeling*. 1998. Disponível em: <http://www.eil.utoronto.ca/enterprise-modelling/papers/>. Acesso em: 21 Novembro 2007.

KIM, H.M.; FOX, M.S.; GRUNINGER, M. *An ontology for quality management; enabling quality problem identification and tracing*. 1999. Disponível em: <http://citeseer.ist.psu.edu/context/1728409/0>. Acesso em: 20 Novembro 2007.

Anexo D: Referências da literatura sobre o método de Uschold e King

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

USCHOLD, M.; KING, M. *Towards a Methodology for Building Ontologies*. 1995. Disponível em: <http://citeseer.ist.psu.edu/uschold95toward.html> Acesso em: 10 Novembro 2007.

USCHOLD, M. *Building Ontologies: Towards a Unified Methodology*. 1996. Disponível em: < http://citeseer.ist.psu.edu/uschold96building.html > Acesso em: 10 Novembro 2007.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.

USCHOLD, M. et al. *Enterprise ontology.* 1998. Disponível em: <http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>. Acesso em: 10 Novembro 2007.

Anexo E: Referências da literatura sobre o método Kactus

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

BERNARAS, A.; LARESGOITI, I.; CORERA, J. *Building and Reusing Ontologies for Electrical Network* Applications. In: PROCEEDINGS OF THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI/96, p. 298-302, 1996.

SCHREIBER, G.; WIELINGA, B; JANSWEIJER, W. *The Kactus View on the 'O' Word.* Workshop on Basic Ontological Issues in Knowledge Sharing/IJCAI95; Montreal, Canada; August 1995. Disponível em: < http://www.cs.vu.nl/~guus/papers/Schreiber95a.pdf > Acesso em: 15 Abril 2008.

Anexo F: Referências da literatura sobre a metodologia Methontology

CORCHO, O.; GÓMEZ-PÉREZ, A.; LÓPEZ-CIMA, A.; LÓPEZ-GARCÍA, V.; SUÁREZ-FIGUEROA, M. C.: ODESeW. *Automatic Generation of Knowledge Portals for Intranets and Extranets*. 2003. Disponível em: <http://webode.dia.fi.upm.es/sew/>. Acesso em: 05 abril 2008.

FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, H. *Methontology:* from ontological art towards ontological engineering. 1997.
Disponível em: <http://citeseer.ist.psu.edu/context/544607/0/>. Acesso em: 20 Agosto 2007.

FERNÁNDEZ, M; GOMEZ-PEREZ, A.; SIERRA, J.P.; SIERRA, A.P . Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, jan./ fev.1999.

FERNÁNDEZ, M. *Overview of methodologies for building ontologies*. 1999. Disponível em: < http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>. Acesso em: 16 Novembro 2007.

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

GÓMEZ-PÉRÉZ, A. *Some ideas and examples to evaluate ontologias*. 1994. Disponível em: <http://citeseer.ist.psu.edu/gomez-perez94some.html>. Acesso em: 12 Setembro 2007.

GÓMEZ-PÉREZ, A., FERNÁNDEZ, M., VICENTE, A.J. *Towards a Method to Conceptualize Domain Ontologies*, ECAI'96 - Workshop on Ontological Engineering, Budapest, August, 1996. Disponível em: <http://citeseer.ist.psu.edu/483876.html>. Acesso em: 12 Novembro 2007.

GÓMEZ-PÉREZ, A. *Evaluation of taxonomic knowledge in ontologies and knowledge bases*. In: WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, 1999, Alberta, Proccedings… Alberta: University of Calgary, 1999. p.6.1.1-6.1.18.

Anexo G: Referências da literatura sobre o método Sensus

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 107-153.

SWARTOUT, B; PATIL, R.; KNIGHT, K.; RUSS, T. *Toward Distributed Use of Large-Scale Ontologies*. 1996.
Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em: 10 Abril 2008.

VALENTE, A.; RUSS, T.; McGREGOR, R.; SWARTOUT, W. Building and (Re) Using an Ontology of Air Campaign Planning. *IEEE Intelligent Systems.* Jan/Fev. 1999. Disponível em: < http://portal.acm.org/citation.cfm?id=630305.630437&coll=GUIDE&dl=GUIDE>. Acesso em: 12 Abril 2008.

Anexo H: Referências da literatura sobre o método 101

NOY, F. N.; GUINNESS, D. L. *Ontology development 101*: a guide to create your first ontology. 2001. Disponível em: <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>. Acesso em: 03 Outubro 2006.

Anexo I: Referências da literatura sobre a metodologia e a norma para construção de vocabulários controlados

ANSI/NISO Z 39.19: 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005. 176 p. Disponível em: <http://www.niso.org/standards/resources/Z39-19-005.pdf?CFID=2436046&CFTOKEN=81298864>. Acesso em: 27 jun. 2007.

CAMPOS, M.L.A.; GOMES, H.E.; MOTTA, D.F. *Tutorial de Tesauro*. 2004. Disponível em: http://www.conexaorio.com/biti/tesauro Acesso em: 30 abril 2008.